An Artificial Intelligence Perspective: How Knowledge and Confidence Shape Risk and Benefit Perception

Nadia Said[1*], Andreea E. Potinteu[1,2] , Irina Brich[2] , Jürgen Buder[2], Hanna Schumm[1], Markus Huff[1,2]

1 Department of Psychology, University of Tübingen; Schleichstraße 4, 72076 Tübingen, Germany.

2 Leibniz-Institut für Wissensmedien; Schleichstraße 6, 72076 Tübingen, Germany.

*Corresponding author. Email: nadia.said@uni-tuebingen.de.

Abstract

Artificial intelligence (AI) applications are increasingly used in everyday life. Whereas some are widely accepted (e.g., automatically compiled playlists), others are highly controversial (e.g., using AI in the classroom). The public discourse reveals a somewhat ambiguous perception of AI, with enthusiasm about the tremendous positive potential of AI on the one hand and growing concerns about the existential dangers of uncontrolled artificial intelligence on the other hand. Thus, understanding the underlying mechanisms of what shapes people's perceptions of AI is vital to help explain AI usage and acceptance trends. In our research, we take a cognitive approach by measuring the perceived risks and benefits of AI applications considering people's knowledge and confidence in their knowledge. To this end, we assessed in two studies (N = 394 and N = 437; representative) how knowledge about AI and confidence in AI knowledge is related to participants' risk-benefit perception of AI scenarios from three domains: media, medicine, and autonomous driving. Results showed that both AI knowledge and confidence in AI knowledge are predictors regarding people's risk-benefit perception beyond people's attitudes toward AI. More specifically, people with more knowledge about AI exhibited so-called risk blindness; that is, they underestimated the risks. On the other hand, higher confidence in one's AI knowledge impacted participants' benefit assessment. Knowledge and confidence thus open a new dimension of understanding people's perception of risks and benefits in AI.

Keywords. artificial intelligence, knowledge, confidence, attitudes, risk-benefit perception, metacognition

Artificial Intelligence (AI) is one of the fastest growing and increasingly popular data-driven technologies used in a wide range of contexts: from media, medicine, transport, criminal justice, finance, education, and even to tackle climate change issues. There is seemingly no end to areas in which AI-based applications can be and are employed, and the usage of AI is currently on the rise: In only four years (2015-2019), the use of AI in organizations rose from 10% to 37% (Gartner, 2019). However, there is not only enthusiasm about this advance of AI in almost all areas of life. A recent US survey (Oliver, 2022), for example, revealed that Americans are rather concerned about the increase in AI-based applications. At the same time, there is a lack of widespread understanding among the public about what AI employs and in which areas it is used (Davenport et al., 2017; Zhang & Dafoe, 2019). While there is no doubt about the rather ambiguous nature of AI – posing a risk and a benefit at the same time – this new technology is here to stay, with possibly long-lasting negative and positive impacts on society. This makes the question of how the public perceives AI-based applications and what factors influence their perception one of utmost importance.

Artificial intelligence applications

The automotive industry is one of the fastest-growing sectors using artificial intelligence applications. For example, vehicles on the road already offer their drivers AI-based assistants to help with parking and lane keeping (Qirko, 2021). Medicine is, likewise, a leading field in terms of AI use, providing personalized treatment plans (Hamet & Tremblay, 2017) and using AI-based diagnostic tools (He et al., 2019) to reduce the risk of undetected illness. Moreover, the

usage of AI-based applications does not stop there: it is used in the entertainment industry, music (AIVA, n.d.), and movies (Rondinella, 2016), and even artwork (Hildebrand, 2015) is now created using the help of AI. Furthermore, fields like quantum mechanics (Duniko & Briegel, 2018), agriculture (Jha et al., 2019), housing (Callaghan et al., 2001), chemistry (de Almeida et al., 2019), education (Zawaki-Richter et al., 2019), sports (Bartlett, 2006), and fashion (Liang et al., 2020) are all benefiting from AI. Usage of AI-based applications is on a steep increase, with estimates suggesting that by 2023, around eight billion digital voice assistants will be in use globally (Virtual Assistant Technology - Statistics & Facts, 2020).

However, as previously mentioned, the ever-growing use of AI applications has sown unease among US citizens (Oliver, 2022). Zhang & Dafoe (2019) also found that 82% of participants expressed that AI should be carefully managed, despite most participants supporting AI use. A similar ambiguous trend can be seen for the Australian population, where AI in healthcare and medicine is seen as positive, while AI use in the workplace is met with doubts (Selwyn et al., 2020). These findings suggest that public perception of AI is rather ambiguous, posing risks and benefits at the same time.

Risks and benefits of AI applications

A large body of literature investigating AI-related risks already exists, showing that perceived risks usually center around the misuse of AI by government or tech companies, loss of privacy, and surveillance (Barth & de Jong, 2017; Kozyreva et al., 2020; Kozyreva et al., 2021; Park & Jones-Jang, 2022; Zhang & Dafoe, 2019; Zhang & Dafoe, 2020). Kelley et al. (2021), for example, conducted a multi-national survey in Australia, Canada, the USA, South Korea, France,

Brazil, India, and Nigeria, corroborating results from previous studies by showing that all countries were expressing concerns over job loss and privacy. Other important factors regarding AI risk perception are age, gender, and education. Siegrist et al. (2005) report that women tend to perceive more risks in some instances, such as technology-related hazards, with older age correlating with higher perceived risk. Other studies reported no difference in age, gender, and education concerning possible data security risks posed by recommendation agents (Rohden & Zeferino, 2022). However, education, age, and gender were identified as essential moderators concerning customers' usage intention of autonomous vehicles (Park & Le, 2021).

Classical risk studies usually focus on topics in which risks are central, such as in man-made nuclear power (Whitfield et al., 2009) or natural disasters (Brown et al., 2018; Lechowska, 2018; Yong & Lemyre, 2019). AI applications as well as new technologies in general (see for example Storer et al. 2023) are ambiguous in nature. That is reflected in the substantial body of literature investigating whether people perceive AI as a benefit (e.g., Antes et al. 2021; Cui & Wu, 2021; Tussyadiah & Miller, 2019; Yeh et al., 2021; with research often also measuring perceived risks alongside). For example, Bao et al. (2022) showed that a myriad of factors, like value dispositions, political and religious ideology, news attention, and topic-specific knowledge influence the perception of AI-related risks and benefits.

Thus, to address this ambiguity about AI in our research, we present participants with scenarios about AI that can be viewed in a *balanced* manner, presenting both the risks and the benefits associated with each scenario. That allows us to investigate how AI-based applications are perceived and, more importantly, whether (meta-)cognitive factors influence this perception,

that is, AI knowledge and confidence in AI knowledge.

Knowledge about AI applications

Research on knowledge about AI consistently showed that the public holds only little knowledge about AI and AI devices (Selwyn et al., 2020; Merenkov et al., 2020). The general public still needs clarification about what AI is and where AI is being used (Edelman, 2019). In a 2019 survey, most respondents could correctly state that terms such as machine learning, robotics, and machine automation are related to AI. However, less than half could correctly state that navigation devices, natural language processing, and computer vision also use AI (Edelman, 2019). Furthermore, the public tends to underestimate AI's prevalence in everyday technology (Zhang & Dafoe, 2019). Currently, the general population's level of knowledge on the topic of AI already is mainly assessed in the form of surveys (Rathi & Rathi, 2019; Selwyn et al., 2020; Selwyn & Cordoba, 2021; Bao et al., 2022), with only a few explicit questions about AI knowledge that are frequently mixed with questions concerned with attitudes towards AI, trust in, and familiarity with AI. A validated AI knowledge test is currently missing in the scientific literature on AI. Thus, the present research aims to develop a reliable AI knowledge test.

Knowledge and risk perception

Despite the large body of research assessing reported knowledge, risk perception, and behavior (Maswanya et al., 1999; Zhang & Norman, 2012; Jones et al., 2012; Iorfa et al., 2020), research on topic-specific knowledge as a predictive factor for ones' risk assessment is less common. Sjöberg and Drottz-Sjöberg (1991), for example, investigated how workers at nuclear

power plants perceive the risks associated with their workplace, with workers assessing risks as higher when they were less informed about potentially detrimental emergency scenarios.

Furthermore, research also shows that experts perceive risks (in their area of expertise) as less severe than laypeople, with an ongoing debate about what drives those differences (Sjöberg, 1998; Sjöberg, 2002; Hansen et al., 2003; Sjöberg, 2008). For example, a study by Gutteling and Kuttschreuter (2011) on the Millennium Bug showed that people with average computer knowledge and no specific expertise were more concerned about the Millennium bug than computer experts. However, our research is not about comparing laypeople to experts but solely about investigating the role of AI knowledge, thus being able to gain some understanding of how levels of AI knowledge among the general public influence their perception of artificial intelligence.

Research investigating the relationship between objective, topic-specific knowledge, and benefit perception is also rather sparse (note, however, that there is substantially more research on general scientific knowledge or science literary and technology endorsement, see, e.g., Brossard & Nisbet, 2007; Nisbet et al., 2002; Peng, 2020). For example, Gigerenzer et al. (2009) investigated whether acquiring knowledge about mammography and PSA screening from different sources was connected to a more accurate perception of the benefits of those procedures. Results showed that, overall, people overestimated the benefits. Thus, the authors concluded that accurate information about the benefits could have been communicated more efficiently to the public. Furthermore, Khare (2023) did find a positive relationship between consumers' knowledge about green apparel and how they perceived the product. Concerning AI-

specific knowledge, Bao et al. (2022) assessed a range of variables among others the level of AI-knowledge as well as risk and benefit perceptions of AI.

However, Bao et al. (2022) do not offer a reliable knowledge scale in their paper. Even more importantly, there seem to be no studies on whether higher knowledge is related to a more *accurate* estimation of risks and benefits. While having more accurate knowledge about a topic could lead to a more accurate risk and benefit assessment, the opposite could also be true. Higher knowledge could also lead people to falsely assume that they are more capable of mitigating any associated risks. That could result in an underestimation of risks, an overestimation of benefits, or both. Some research is already hinting towards a relationship between higher levels of expertise and the subjective feeling of being more in control (Skjong & Wentworth, 2001; Du et al., 2007; Cho et al., 2010; Dikmen et al., 2018).

Thus, a reasonable expectation would be that higher AI knowledge could (mis)lead people to assume that benefits related to AI outweigh the accompanying risks. As such, we expect participants with higher AI knowledge to view AI-based applications more as a benefit than a risk, exhibiting *risk blindness*.

Confidence in AI knowledge

Being (over)confident about one's knowledge can have serious consequences affecting people's judgments and decision-making (Ahmad, 2020; Cassam, 2017; Malmendier & Tate, 2015; Mannes & Moore, 2013; McKenzie et al., 2008; Naguibet al., 2019; Pikulina et al., 2017; Yang et al., 2012). For example, Canfield at al. (2019) showed that participants' exhibited a metacognitive bias, that is they were overconfident in their ability to correct identify phishing

mails (see also Li & Rao, 2022; Sarno & Neider,2022; Sarno et al., 2022). Other research showed that participants with IT background were overconfident in their ability to detect deep fakes (Sütterlin et al., 2022). Pengnate and Antonenko (2013), investigated whether metacognitive awareness plays a moderating role regarding participants' website trustworthyness ratings. Results showed that metacognitive awareness is no related to perceived website trustworthyness.

Research has already related overconfidence, in general, to risky behavior (Broihanne et al., 2014; Robinson & Marino, 2015; Nosić & Weber, 2010; Glaser & Weber, 2007). However, studies relating *confidence in one's factual knowledge* about a topic to risk or benefit perception of the said topic to our knowledge do not currently exist. Nevertheless, at least one study shows a positive relationship between people's confidence in their knowledge (measured by asking them how they subjectively would estimate their knowledge about a specific topic) and a more positive view of technology (Wildavsky & Dake, 1990). Furthermore, some studies have investigated how people's subjective ratings of their knowledge related to their risk perception and behavior, providing mixed evidence. Jaccard et al. (2005) found that individuals who rated their knowledge as high tended to show riskier behavior regarding contraceptive use.

Conversely, in a study by Zhu et al. (2016), individuals who perceived their knowledge level as high rated the risk of nuclear energy usage as higher than those who rated their knowledge level lower. Given these findings, one could expect a relationship between confidence in knowledge and risk perception. However, the direction of this relationship is unclear. Thus, instead of a directional influence, we expect high confidence to lead to a higher

difference between the risk and benefit ratings. We anticipate that people who are confident that their knowledge of AI is correct would make them biased and push them towards either overestimating the risks or the benefits, regardless of their actual knowledge of AI.

The current research

Given the staggering rise of AI applications, it is evident that technology based on AI will be with us in the years to come, making the issue of how the public perceives AI-based applications and what factors influence their perception more relevant than ever. In our research, we aim to gain a deeper understanding of people's risk-benefit perception of AI while assessing AI-related knowledge and confidence in AI knowledge in two studies. Thus, the aims of our current paper are twofold: (i) creating a reliable test for AI-related knowledge and (ii) bridging the literature gaps highlighted above: understanding the relationship between the risk-benefit assessments for AI and AI knowledge and confidence in AI knowledge. Based on the existing literature, we expect to see an influence of AI knowledge on the strength and direction of risk and benefit assessments. We hypothesize that an elevated level of knowledge leads knowledgeable people to rate the benefits in AI higher than the risks. Additionally, we expect confidence in AI knowledge to influence the differences between the risk and benefit ratings, with high confidence leading to a high difference between the ratings.

Study 1

Participants

The sample size for the study was calculated based on the approximate population size of Germany (80 million), with a confidence level of 95% and an error margin of 5% (Rahi et al., 2019; Cochran, 1977), resulting in N = 385 participants[1]. To account for data exclusions based on the per-registered criteria (i.e., failed attention check), we aimed at N = 405 participants (i.e., oversampling of 5%). The total number of participants who completed the survey was N = 408. After excluding participants who failed the attention check, the final sample consisted of N = 394 participants, all of whom fully completed the questionnaires and consented to participate in the study. The online platform *Prolific* was used to recruit the participants. Participants were monetarily compensated for participating in the study at a rate of 2.50 Pounds per 20 minutes. The mean age of participants was M = 27.45 years (SD= 9.27). The sample consisted of 228 male participants and 160 female participants. Furthermore, five participants reported their gender as nonbinary, and 1 participant chose “not specified”[2]. In terms of education, 4.6 % of the sample reported having graduated from high school (Volks-, Hauptschul- oder gleichwertiger Abschluss); 17% reported having graduated from a secondary school (Mittlere Reife,

1 The sample size was calculated based on the following formula: $\frac{\frac{z^2 \times p(1-p)}{e^2}}{1+\frac{z^2 \times p(1-p)}{e^2 N_{population}}}$ with $N_{population}$ = population size, e = error margin, z = Z-Value = 1.96 for a confidence level of 95%.

2 Because of the very low number of participants in the gender categories “nonbinary” (n = 5) and “not specified” (n = 1), those were randomly assigned to the categories “male” or “female” to keep them in the regression analysis. Furthermore, participants who chose “other” in the education category (n = 5) were excluded from the regression analysis to be able to assign numbers to the different education levels, making them continuous.

Realschulabschluss- oder gleichwertiger Abschluss), 6,9% reported having completed an apprenticeship (Abgeschlossene Lehre), 29.7% reported having completed their college level diploma ((Fach-)Hochschulreife, (Fach-)Abitur), 39.8% reported having graduated from a university (abgeschlossenes (Fachhochschul-)Studium), 0.8% of participants reported having no degree and 1.3% of them reported having finished some other education.

Procedure

Participants were informed about the study and agreed to the requirements (being over 18 years old and fluent in German) to proceed further. Afterward, demographic data were collected, including age, gender, and education level.

The main study was completed in the following order: First, participants were asked to give a subjective assessment of their knowledge on the topic of AI on a five-point scale, ranging from "very low" to "very high". Second, participants had to answer the AI knowledge questionnaire and indicate their confidence in their answer for each knowledge item. Third, participants read one AI scenario and gave their risk-benefit assessment regarding the scenario. In the last step, participants had to give their consent, agreeing that their data would be evaluated for the purposes of the study.

Measures

AI knowledge

To measure participants' *AI knowledge* objectively, a questionnaire was developed. The items covered nine domains: functioning, medicine, autonomous driving, education, personnel recruiting, media, communication, environment, and history. The development of the

questionnaire was part of a BSc thesis (Schumm, H., 2021). The nine domains were chosen out of a larger pool of possible areas of application based on an extensive literature search. For the medical domain for example the most important as well as most common AI applications were chosen. That is, AI could be used as diagnostic tools (Yala et al., 2019; He et al., 2019), for treatment suggestions (Hamet & Trembley, 2017), or as support for patient care and nursing (Janowski et al., 2018). The questionnaire consisted of forty-six items (twenty-eight true/eighteen false items) with a dichotomous response scale (true/false). The original items in German and an English translation are in Table A1 in the appendix.

Confidence in AI knowledge

To measure *confidence* in AI knowledge, participants were asked to indicate their confidence in their answers for each knowledge item. Participants could answer on a 6-point percentage scale, from “I guessed – 50%” to “I am sure – 100%”. The intermediary percentages were given as exclusively numeric (e.g., 60%), with no additional linguistic descriptions.

Perception of artificial intelligence

To measure participants' *perception of artificial intelligence*, participants were randomly presented with one out of nine *scenarios* covering three domains: medical domain, autonomous driving, and media. These domains were chosen as the most likely to be encountered in day-to-day life and, as such, presumably easily accessible to the public. In the development of the scenarios, we focused on presenting an everyday situation in which the participant might encounter AI. Additionally, we focused on developing *balanced scenarios* regarding the risks and benefits associated with each. The length of the scenarios described ranged from 80 to 151

words. Furthermore, they were all written in a first-person perspective to increase relatability. For example, one of the medical scenarios was described as follows: "You discover a spot on your skin that is getting bigger and bigger. Worried that it is skin cancer, you take a picture of the spot and upload it to a website where it is analyzed by artificial intelligence. After a short wait, the AI concludes that it is not malignant skin cancer. The AI informs you of this and tells you not to worry. As a result, you decide not to make a doctor's appointment."

The original scenarios in German and the English translations can are in Table A2 in the appendix.

Risk-benefit assessment

Risk assessment of the scenarios was done using an adapted version of the questionnaire proposed by Walpole et al. (2021). The original questionnaire comprised four categories: affect, exposure, susceptibility, and severity. Due to an implementation error, only two categories, affect and susceptibility, were used in the analysis. Participants rated each risk item on a 5-point Likert scale. For example, the options for the question "How concerned are you, if at all, about this AI application?" were: "not concerned at all", "slightly concerned", "moderately concerned", "very concerned", "extremely concerned" (all questions and answer options are in Table A3 in the appendix).

Benefit assessment was done by adapting the items from the Walpole et al. (2021) questionnaire by reversing the options used for the risk items by replacing, for example, "concerned" with "optimistic". Participants rated each item on a 5-point Likert scale. For example, the options for the question "How optimistic are you, if at all, about this AI

application?" were "not optimistic at all", "slightly optimistic", "moderately optimistic", "very optimistic", "extremely optimistic" (all questions and answer options are in Table A3 in the appendix).

Analysis

AI knowledge

Participants' *knowledge accuracy* was calculated by averaging the participants' responses to the knowledge items. Additionally, we determined participants' *knowledge sensitivity* d' as specified in a Signal Detection Theory (SDT) framework by calculating the difference of the Z (True Positive) – Z (False Positive) rate, with Z being the inverse cumulative density function of the normal distribution. This approach has the advantage of factoring out the participants' response biases (tendency to answer "yes") (Macmillan & Creelman, 2004). This is especially important when the number of true and false items differ considerably (Study 1: 28 true and 18 false items, Study 2: 28 true and 24 false). Thus, knowledge sensitivity is a bias-free measure of the participant's knowledge accuracy. For both studies, analyses were run twice: once with the knowledge score and once with knowledge sensitivity. As there was no difference regarding the results for *all calculations,* only the results for the knowledge score are reported in the following. Note that the R code includes both scores and calculations.

Confidence in AI knowledge

Participants' *confidence score* was calculated by averaging participants' confidence responses.

Risk-benefit assessment

Participants' *risk score* was calculated by averaging participants' answers to the risk items. Participants' *benefit score* was calculated by averaging participants' answers to the benefit items. Due to implementation issues, the susceptibility items had to be excluded from the calculation. Furthermore, we separately calculated Cronbach's alpha for the risk and benefit items. Cronbach's alphas for the nine risk and nine benefit items were .66 and .82, respectively. To increase reliability, we excluded the exposure items resulting in an $\alpha = .82$ for the six remaining risk items and $\alpha = .87$ for the six remaining benefit items (based on *the psych::alpha* package in R). Those items were used to compute the two *risk-benefit scores*, RB1 and RB2: RB1 was calculated as the difference between the benefit and the risk score. Thus, a positive sign denotes that the benefits outweigh the risks, while a negative sign denotes that the risks outweigh the benefits. RB2 was calculated as the absolute value of the differences between the risk and benefit scale scores. Thus, smaller values denote a more balanced view of risks and benefits, while higher values denote a bigger difference between risk and benefit perception. While the RB1 score allows us to investigate the *strength and direction* of the risk and benefit assessment, the RB2 score sheds light on the *difference* between the two scores. For all analyses, we always also investigated the relationship between AI knowledge and confidence in AI knowledge separately for perceived risks and benefits. This was done to gain further insight into the driving factor/s behind the results: perceived risks, benefits, or both.

Effect sizes. For all analyses we calculated effect sizes using Cohen's f^2. Effect sizes of $f^2 \geq .02$ are interpreted as small, effect sizes of $f^2 \geq .15$ are interpreted as medium, and effect sizes

of $f^2 \geq .35$ are interpreted as large (Cohen, 2013).Values for the confidence intervals for f^2 can only be >= 0. As the output only shows two decimal places and given that the confidence intervals were determined by bootstrapping, whether or not the effect sizes are significant are difficult to tell in cases in which the confidence intervals contain a zero on the left side (which is the case in some of our calculations). Note however, that in all cases in which adding confidence and/or knowledge to the baseline model was significant, model comparison showed a significant ΔR^2 (see respective Tables in the appendix).

Results

Descriptive results

Means, standard deviations, and correlations of the main variables are displayed in Table 1.

Table1

Means, standard deviations, and correlations of the main variables with confidence intervals

Variable	*M*	*SD*	1	2	3	4	5	6
Risk	2.50	0.77						
Benefit	2.73	0.86	-.42** [-.50, -.34]					
Knowledge	0.65	0.08	-.11* [-.20, -.01]	-.04 [-.14, .06]				
d-prime	0.90	0.50	-.11* [-.21, -.01]	-.02 [-.11, .08]	.95** [.94, .96]			
Confidence	3.53	0.74	-.03 [-.13, .07]	.15** [.05, .24]	.19** [.10, .29]	.23** [.13, .32]		
RB1 score	0.23	1.38	-.82** [-.85, -.79]	.86** [.83, .88]	.03 [-.07, .13]	.05 [-.05, .15]	.11* [.01, .21]	

RB2 score	1.11	0.85	.07	.05	.18**	.18**	.14**	-.00
			[-.03, .16]	[-.05, .15]	[.08, .27]	[.08, .27]	[.04, .23]	[-.10, .09]

Note. M and *SD* represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$, ** indicates $p < .01$.

AI knowledge

Overall participants answered 65% (*SD* = 8.11%) of the items correctly. Cronbach's alpha for the knowledge items was .51.

Risk-benefit assessment

Risk and benefit ratings did not differ for scenarios 1, 2, 5, 7, and 8 (gray). Scenarios 3, 4, and 6 were rated as a benefit, and scenario 9 was rated more as a risk than a benefit (see Table 2 for t-test results). Note that the results were Bonferroni corrected with an alpha level of .005. Overall, scenarios were rated rather as a benefit (*M* = 2.70, *SD* = 0.68) than a risk (*M* = 2.45, *SD* = 0.75), t (416.64) = -3.42, $p < .001$ (see Figure 1).

Figure 1

The figure displays the average risk-benefit assessment, separately for risks (magenta) and benefits (purple) for the three different domains: medical domain (left), autonomous driving (middle), and media (right).

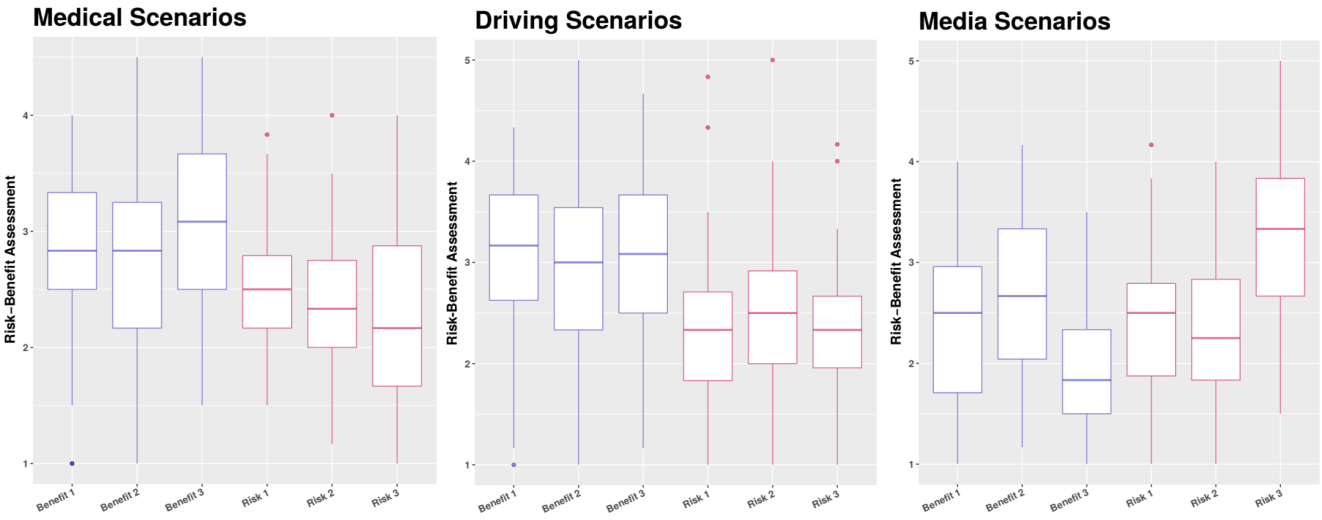


Table 2

Means, standard deviations, and t-test results for the nine different scenarios

Nr.	Scenarios	M_{risk}	SD_{risk}	$M_{benefit}$	$SD_{benefit}$	*T*	P
Scenario 1	Medical 1	2.51	0.73	2.78	0.69	t(78.603) = -1.92	.058
Scenario 2	Medical 2	2.34	0.73	2.67	0.66	t(76.914) = -1.99	.050
Scenario 3	Medical 3	2.33	0.71	3.07	0.72	t(85.885) = -4.38	< .001
Scenario 4	Driving 1	2.32	0.73	3.01	0.65	t(85.870) = -4.09	< .001
Scenario 5	Driving 2	2.57	0.74	2.96	0.56	t(83.264) = -2.15	.035
Scenario 6	Driving 3	2.34	0.79	3.04	0.79	t(83.616) = -4.64	< .001
Scenario 7	Media 1	2.44	0.68	2.39	0.77	t(81.954) =0.31	0.754
Scenario 8	Media 2	2.37	0.83	2.67	0.73	t(86.389) =-1.94	.056
Scenario 9	Media 3	3.23	0.83	1.98	0.72	t(86.514) =7.38	< .001

Influence of knowledge: Does knowledge influence the strength and direction of the risk and benefit assessment?

Multiple linear regressions were performed for the risk-benefit score (RB1) to assess whether AI knowledge influences the strength and direction of the risk and benefit assessment while controlling for confidence, gender, age, and education (see Table B1 in the appendix).

Results showed that the predictors explained a significant amount of variance in the RB1 score, $F(5, 383) = 4.25$, $p < .001$, $R^2 = .05$, $R^2_{Adjusted} = .04$, with age, b = -.03, $t(383) = -3.56$, $p < .001$, Cohen's $f^2 = .03$, 95% CI [.01, .08], significantly predicting the RB1 score. For knowledge, however, we did not find a significant influence b = .38, $t(383) = .44$, $p = .66$. Thus, for younger participants, benefits of AI outweighed the risks.

Influence of confidence: Does confidence influence the differences between the risk and benefit ratings?

Multiple linear regression was performed for the risk-benefit score (RB2) to assess whether confidence in AI knowledge influences the differences between the risk and benefit ratings while controlling for AI knowledge, gender, age, and education (see Table B2 in the appendix). Results indicated that the investigated predictors explained a significant amount of variance in the RB2 score, $F(5, 383) = 5.32$, $p < .001$, $R^2 = .06$, $R^2_{Adjusted} = .05$. The analysis showed that gender, b = .22, $t(383) = 2.46$, $p = .014$, Cohen's $f^2 = .00$, 95% CI [.00, .03], knowledge, b = 1.65, $t(383) = 3.09$, $p = .002$, Cohen's $f^2 = .03$, 95% CI [.01, .08], and confidence, b = .15, $t(383) = 2.56$, $p = .011$, Cohen's $f^2 = .02$, 95% CI [.00, .05], significantly predicted the RB2 score. That is, higher confidence in AI knowledge and a higher knowledge level was related to a larger difference between risk and benefit ratings.

Risk or benefit? Separate assessment of the risk and benefit scores

Multiple multivariate regressions were conducted to evaluate the influence of AI knowledge on the risk and benefit scores separately, controlling for confidence, gender, age, and education. Results show that for the *benefit scores*, the investigated predictors explained a

significant amount of variance, $F(5, 383) = 5.73$, $p < .001$, $R^2 = .07$, $R^2_{Adjusted} = .06$. Additionally, age, b = -.02, $t(383) = -3.64$, $p < .001$, Cohen's $f^2 = .04$, 95% CI [.01, .09] and confidence, b = .20, $t(383) = 3.27$, $p = .0011$, Cohen's $f^2 = .03$, 95% CI [.00, .07], significantly predicted the benefit scores (see Table B3, in appendix). That is, younger participants and participants with higher confidence in their AI knowledge had a higher benefit score.

For the *risk scores,* results showed that the investigated predictors did explain a significant amount of variance, $F(5, 383) = 2.62$, $p = .024$, $R^2 = .03$, $R^2_{Adjusted} = .02$. Additionally, the predictors age, b = .01, $t(383) = 2.30$, $p = .022$, Cohen's $f^2 = .01$, 95% CI [.00, .04], and knowledge, b = -1.05, $t(383) = -2.13$, $p = .034$, Cohen's $f^2 = .01$, 95% CI [.00, .04], significantly predicted the risk scores (see Table B4, in appendix). That is, older participants and participants with lower AI knowledge had a higher risk score.

Summary Study 1

The first study investigated the influence of knowledge and confidence on the participants' risk-benefit perception for different AI scenarios. Results showed that only age was predictive for our first measurement, RB1. That is, for younger participants, benefits outweighed the risks. For our second measurement, RB2, not only confidence but – surprisingly – also knowledge were positively associated with a greater difference between the risk-benefit ratings. Furthermore, separate analyses for the risk and the benefit scores revealed that being older and having a lower AI knowledge score was associated with higher risk scores (note that effect sizes in both cases were very small) while being younger and having higher confidence in AI knowledge was associated with higher benefit scores.

Study 1 has several limitations: First, the AI knowledge questionnaire had an alpha of only .51. This might be a likely reason why we did not find any relationship between knowledge and our first risk-benefit measurement (RB1). Second, to assess whether confidence and knowledge influence risk-benefit perception, scenarios must be, on average, rated equally regarding their risks and benefits. However, even though the difference was small, scenarios were mostly rated as more of a benefit than a risk. Third, differences in the participants' attitudes toward AI might also be an essential impact factor (Araujo et al., 2020; Zhang & Dafoe, 2019). Thus, it is vital to assess whether the influence of confidence and knowledge is a more significant influencing factor than people's attitudes toward AI. In Study 2, we addressed these limitations by (i) improving the existing AI knowledge test, (ii) adapting the AI scenarios to ensure a more balanced risk-benefit assessment, and (iii) including an attitude questionnaire. Furthermore, we opted to recruit a representative German sample to ensure that our results can be reliably generalized to the German population.

Study 2

Participants

As in the first study, the sample size for the second study was also calculated based on the approximate population size of Germany (80 million), together with a confidence level of 95% and an error margin of 5% (Rahi et al., 2019; Cochran, 1977) resulting in $N = 385$ participants. To account for data exclusions based on the pre-registered criteria (i.e., failed attention check), we aimed at $N = 405$ participants (i.e., oversampling of 5%). The total number

of participants who completed the survey was *N*= 454. After excluding participants who failed the attention check, as well as those who did withdraw their consent for data processing at the end of the study, the final sample consisted of *N* = 437 participants, all of whom fully completed the questionnaires and gave their consent to be a part of the study. The online platform *respondi* was used to recruit a *representative sample* of the German population. Participants were monetarily compensated for taking part in the study with a rate of 3.60 Euros per 20 minutes. The mean age of participants was *M* = 44.5 years (*SD* = 14.8). The sample consisted of 210 male participants and 224 female participants. Furthermore, 2 participants reported their gender as nonbinary, and 1 participant chose "not specified"[3]. In terms of education, 5.9% of the sample reported having graduated high school (Volks-, Hauptschul- oder gleichwertiger Abschluss); 25.8% reported having graduated from a secondary school (Mittlere Reife, Realschulabschluss- oder gleichwertiger Abschluss), 18% reported having completed an apprenticeship (Abgeschlossene Lehre), 24,7% reported having completed their college level diploma ((Fach-)Hochschulreife, (Fach-)Abitur), 24,9% reported having graduated from a university (abgeschlossenes (Fachhochschul-)Studium), 0.2% of participants reported having no degree and 0.2% of them reported having finished some other education.

Procedure

3 Because of the very low number of participants in the gender categories "nonbinary" (n = 2) and "not specified" (n = 1) those were randomly assigned to the categories "male" or "female" to keep them in the regression analysis. Furthermore, participants who chose "other" in the education category (n = 1) were excluded from the regression analysis to be able to assign numbers to the different education levels making them continuous.

The procedure for Study 2 was identical to Study 1 with one exception: before completing the AI knowledge test, participants were now also asked to indicate their attitudes towards AI.

Measures

The measures were identical to those in Study 1 with three exceptions: (i) we updated the AI knowledge items, (ii) we updated the AI scenarios, and (iii) we included items assessing the participants' attitudes towards AI.

AI knowledge

AI knowledge was measured using an adapted version of the previously developed questionnaire. We extended the questionnaire by adding 17 new false items, resulting in a total of 28 true and 24 false items (see appendix Table A1)[4].

Perception of artificial intelligence

To measure the participants' *perception of artificial intelligence* participants were randomly presented with one out of six *scenarios* covering three domains: medicine, autonomous driving, and media. Scenarios were adapted versions of the ones used in Study 1. We selected scenarios noted with 1, 2, 7, and 8 in Table A2, as their risks and benefit ratings did not, or only marginally differ. To have an equal number of scenarios within each domain, we also included scenarios 5 and 4. Scenarios 2, 4, 5, and 8 were adapted to ensure a more balanced risk-benefits assessment (see Appendix Table A2).

4 Note that due to an implementation error, the six items kept from Study 1 were actually those leading to a decrease of alpha in Study 1. Therefore, they were excluded from the analysis in Study 2.

Risk-benefit assessment

Risk and benefit assessment of the scenarios were again done using an adapted version of the questionnaire proposed by Walpole et al. (2021). To be consistent with Study 1, we only included the affect and susceptibility categories as measures.

Attitudes towards AI

Attitudes were measured using a 5-item attitude questionnaire (Sindermann et al., 2020): 1. I fear artificial intelligence; 2. I trust artificial intelligence; 3. Artificial intelligence will destroy humankind; 4. Artificial intelligence will benefit humankind; 5. Artificial intelligence will cause many job losses. Answers ranged from “strongly disagree” to “strongly agree”, on an 11-point scale. Items were re-coded in such a way that high values denoted a more positive attitude towards AI.

Analysis

The analysis in Study 2 was identical to the analysis in Study 1.

Results

Descriptive results

Means, standard deviations, and correlations of the main variables are displayed in Table 3.

Table 3

Means, standard deviations, and correlations of the main variables with confidence intervals

Variable	*M*	*SD*	1	2	3	4	5	6	7
Risk	2.74	0.89							
Benefit	2.52	0.87	-.18** [-.27, -.09]						

Knowledge	0.60	0.11	-.24** [-.33, -.15]	.06 [-.04, .15]					
d-prime	0.55	0.66	-.24** [-.33, -.15]	.07 [-.02, .16]	.97** [.97, .98]				
Confidence	2.83	1.14	-.12* [-.21, -.03]	.13** [.03, .22]	.33** [.24, .41]	.34** [.26, .42]			
RB1 Score	-0.21	1.35	-.78** [-.81, -.74]	.76** [.72, .80]	.19** [.10, .28]	.21** [.11, .29]	.16** [.07, .25]		
RB2 Score	0.99	0.95	.21** [.12, .30]	-.23** [-.32, -.14]	.08 [-.02, .17]	.06 [-.03, .15]	.15** [.06, .24]	-.28** [-.37, -.20]	
Attitudes	-5.64	1.94	-.39** [-.46, -.31]	.29** [.20, .37]	.26** [.17, .34]	.27** [.18, .35]	.28** [.19, .37]	.44** [.36, .51]	.02 [-.11, .08]

Note. M and *SD* represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$, ** indicates $p < .01$.

AI knowledge

On average, participants correctly answered about 60% (*SD* = 11%) of the 46 questions used in total. Cronbach's alpha for the knowledge items was α = .68. The final items are marked with an "**" and "***" in Table A1.

Risk-benefit assessment

Risk and benefit ratings did not differ for scenarios 1, 2, 3, 4, and 6 (gray). Scenario 5 was rated as risk rather than benefit (see Table 7 for t-test results). Note that results were Bonferroni corrected with an alpha level of .008. Overall, scenarios were rated rather as a risk (*M*

= 2.73, *SD* = 0.67) than a benefit (*M* = 2.52, *SD* = 0.65), *t*(871.37) = 3.59, *p* < .001 (see Figure 2).

Figure 2

The figure displays the average risk-benefit assessment, separately for risks (magenta) and benefits (purple) for the three different domains: medical domain (left), autonomous driving (middle), and media (right).

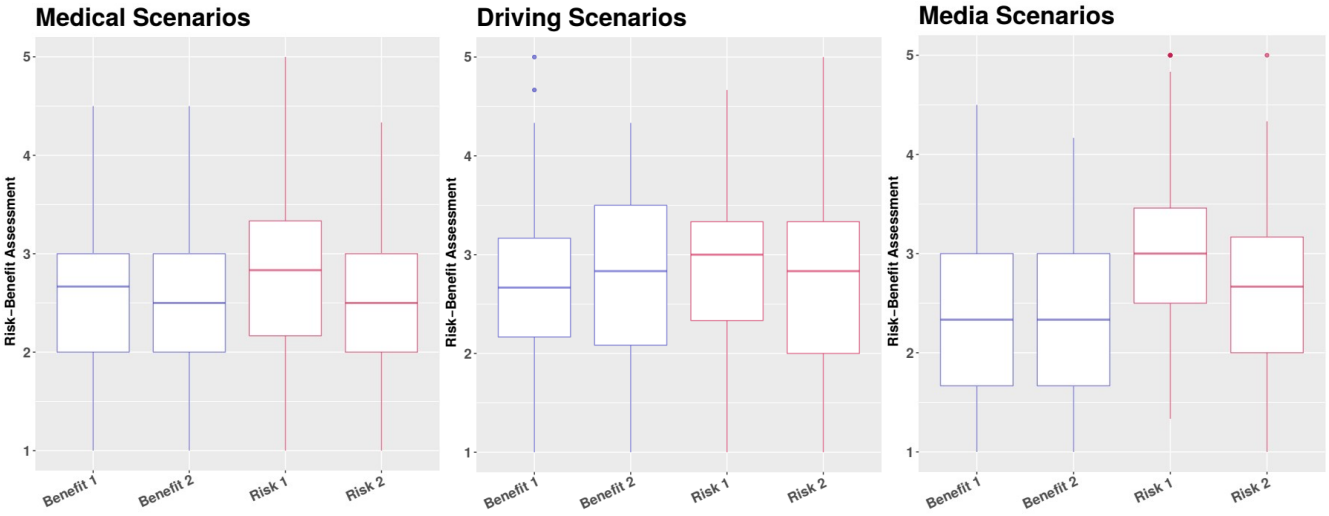


Table 4

Means, standard deviations, and t-test results for the six different scenarios

Nr.	Scenarios	M_{risk}	SD_{risk}	$M_{benefit}$	$SD_{benefit}$	*t*	*P*
Scenario 1	Medical 1	2.75	0.72	2.56	0.61	t(146) = 1.28	.201
Scenario 2	Medical 2	2.46	0.63	2.53	0.56	t(139.95) = -0.47	0.638
Scenario 3	Driving 1	2.81	0.61	1.65	0.65	t(145.82) = 1.00	.316
Scenario 4	Driving 2	2.74	0.63	2.74	0.74	t(147.91) = -0.02	0.976
Scenario 5	Media 1	3.03	0.68	2.34	0.72	t(136.23) = 4.73	>.001
Scenario 6	Media 2	2.63	0.75	2.29	0.64	t(144) = 2.42	.016

Influence of knowledge: Does knowledge influence the strength and direction of the risk and benefit assessment?

Multiple linear regressions were performed on the risk-benefit assessment (RB1) score to assess whether knowledge influences the strength and direction of the risk and benefit assessment while controlling for confidence, gender, age, and education (see Table C1 in the

appendix). Results showed that a significant amount of variance was explained by the investigated variables, $F(5, 430) = 6.48$, $p < .001$, $R^2 = .07$, $R^2_{Adjusted} = .05$. Additionally, age, b = -.01, $t(430) = -2.06$, $p = .04$, Cohen's $f^2 = .00$, 95% CI [.00, .03], gender, b = -.34, $t(430) = -2.57$, $p = .01$, Cohen's $f^2 = .03$, 95% CI [.00, .07], and knowledge, b = 1.96, $t(430) = 3.20$, $p = .0015$, Cohen's $f^2 = .02$, 95% CI [.00, .06], were significant predictors for the RB1 score.

Influence of confidence: Does confidence influence the differences between the risk and benefit ratings?

We conducted multiple regression analysis for the risk-benefit score (RB2) to assess whether confidence in AI knowledge influences the differences between the risk and benefit ratings while controlling for AI knowledge, gender, age, and education (see Table C2, in the appendix). Results showed that the investigated predictors explained a significant amount of variance, $F(5, 430) = 4.81$, $p < .001$, $R^2 = .05$, $R^2_{Adjusted} = .04$. Additionally, age, b = .01, $t(430) = 3.58$, $p < .001$, Cohen's $f^2 = .03$, 95% CI [.01, .07], and confidence, b = .12, $t(430) = 2.88$, $p < .004$, Cohen's $f^2 = .02$, 95% CI [.00, .05], significantly predicted the RB2 score.

Risk or benefit? Separate assessment of the risk and benefit scores.

We performed multiple regression analyses on the risk and benefit scores separately, with knowledge, confidence, age, gender, and education as predictors. Results showed that the investigated variables explained a significant amount of variance in the *benefit score*, $F(5, 430) = 5.37$, $p < .001$, $R^2 = .06$, $R^2_{Adjusted} = .04$, with age, b = -.01, $t(430) = -4.07$, $p < .001$, Cohen's $f^2 = .03$, 95% CI [.01, .08], and gender, b = -.17, $t(430) = -1.99$, $p = .047$, Cohen's $f^2 = .02$, 95% CI [.00, .08], being significant predictors for the benefit scores. The confidence results were only

marginally significant: b = .08, $t(430) = 1.94$, $p = .053$ (see Table C3, in the appendix).

When analyzing the *risk ratings* by means of multiple linear regression, with education, age, gender, confidence, and knowledge as predictors, results indicated that a significant amount of variance was explained by the predictors, $F(5, 430) = 7.03$, $p < .001$, $R^2 = .07$, $R^2_{Adjusted} = .06$. Additionally, knowledge, b = -1.68, $t(430) = -4.19$, $p < .001$, Cohen's $f^2 = .04$, 95% CI [.01, .09], significantly predicted the risk score (see Table C4, in the appendix).

Influence of attitudes towards AI

To analyze whether the influence of knowledge and confidence prevails beyond participants' attitudes regarding AI, we conducted multiple regression analyses controlling for education, age, gender, and attitudes.

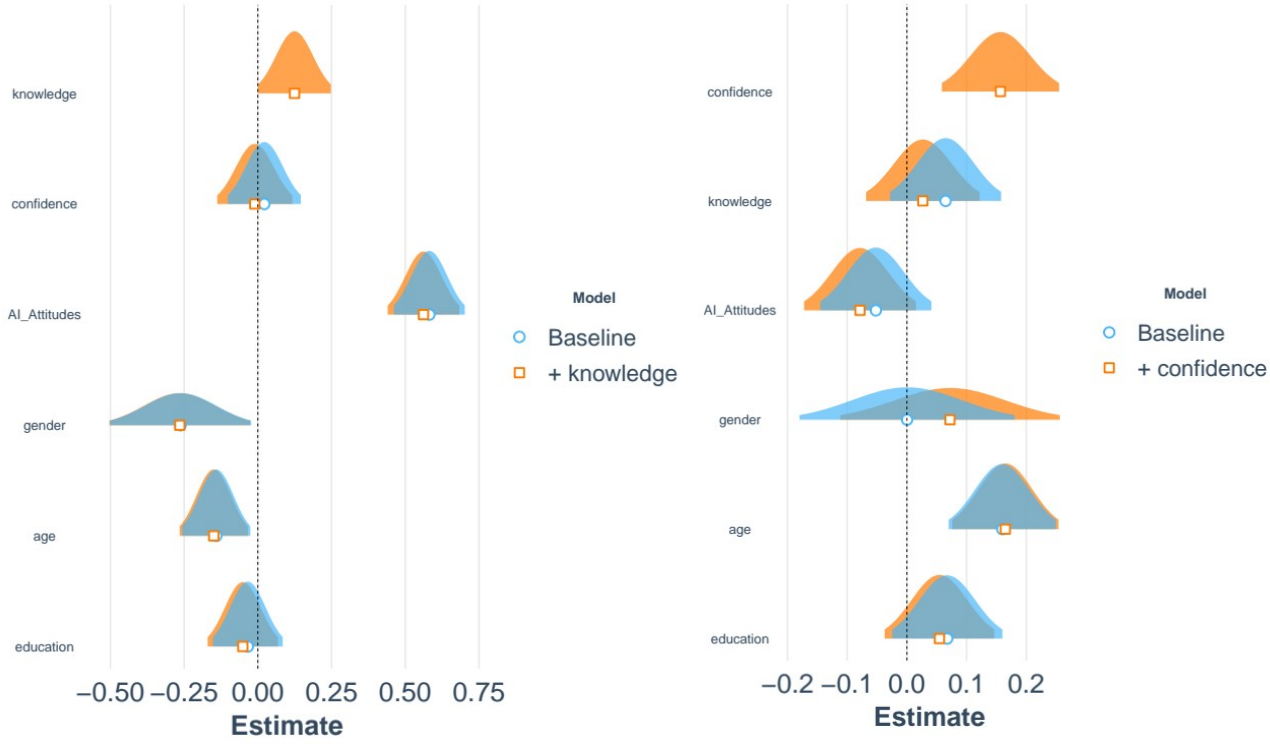
The results indicate that a significant amount of variance in the RB1 score was explained by the predictors, $F(6, 429) = 20.18$, $p < .001$, $R^2 = .02$, $R^2_{Adjusted} = .02$. Furthermore, age, b = -.01, $t(429) = -2.54$, $p = .01$, Cohen's $f^2 = .00$, 95% CI [.00, .03], gender, b = -.26, $t(429) = -2.18$, $p = .03$, Cohen's $f^2 = .03$, 95% CI [.01, .07], knowledge, b = 1.12, $t(429) = 1.98$, $p = .048$, Cohen's $f^2 = .00$, 95% CI [.00, .04], and attitudes, b = .29, $t(429) = 9.09$, $p < .01$, Cohen's $f^2 = .23$, 95% CI [.14, .33], significantly predicted the RB1 score (see Figure 3 (left panel) & Table C5, in the appendix).

An additional multiple regression was conducted for the RB2 score, with predictors explaining a significant amount of variance, $F(6, 429) = 4.48$, $p < .001$, $R^2 = .06$, $R^2_{Adjusted} = .04$. Here, age, b = .01, $t(429) = 3.64$, $p < .001$, Cohen's $f^2 = .03$, 95% CI [.01, .07], and confidence, b = .14, $t(429) = 3.14$, $p < .0018$, Cohen's $f^2 = .02$, 95% CI [.00, .06], were significant predictors

for the RB2 score (see Figure 3, right panel, and Table C6, in the appendix).

Figure 3

Influence of knowledge and confidence on risk-benefit assessment



Note. The figure displays the results of multiple regression models predicting the risk-benefit score RB1 and RB2 with age, gender, education, and attitudes towards AI as control variables. Left: Regression coefficients from the linear regression models predicting RB1 score. The size of the distribution indicates 95% CI. Right: Regression coefficients from the linear regression models predicting RB2 score. The size of the distribution indicates 95% CI.

Introducing attitudes as a predictor for the risk and benefit ratings separately provides further insight. We ran a multiple regression to investigate the influence of attitudes on the *benefit score* while controlling for knowledge, confidence, gender, age, and education. Results showed that the investigated variables explained a significant amount of variance in the benefit

score, $F(6, 429) = 10.83$, $p < .001$, $R^2 = .01$, $R^2_{Adjusted} = .01$. The results also indicated that education, b = -.07, $t(429) = -1.99$, $p = .048$, Cohen's $f^2 = .00$, 95% CI [.00, .00], age, b = -.01, $t(429) = -4.43$, $p < .001$, Cohen's $f^2 = .04$, 95% CI [.01, .08], and attitudes, b = .13, $t(429) = 6.00$, $p < .001$, Cohen's $f^2 = .10$, 95% CI [.05, .16], were significant predictors for the benefit score (Table C7, in the appendix). We ran the same analysis for the *risk ratings.* Results showed again that a significant amount of variance was explained by the investigated predictors, $F(6, 429) = 16.05$, $p < .001$, $R^2 = .02$, $R^2_{Adjusted} = .02$. Furthermore, results indicated that knowledge, b = -1.22, $t(429) = -3.19$, $p = .0015$, Cohen's $f^2 = .02$, 95% CI [.00, .06], and attitudes, b = -.16, $t(429) = -7.45$, $p < .001$, Cohen's $f^2 = .16$, 95% CI [.09, .24], were significant predictors for the risk score (Table C8, in the appendix).

Summary Study 2

The purpose of the second study was to (i) improve the AI knowledge test as well as the scenarios, (ii) investigate the influence of attitudes towards AI in addition to knowledge and confidence on risk-benefit perception, and (iii) assess whether our results from Study 1 would be replicable within a representative sample. Results showed that: first, the alpha of our knowledge test improved with the new items ($\alpha = 0.68$) and second, results for our RB1 and RB2 scores were now in line with our hypotheses: That is, higher knowledge leads to a more positive perception of AI, that is benefits outweigh the risks (RO1score), while higher confidence leads to a higher difference between the risk and benefit ratings (RB2 score). When investigating risks and benefits separately, we largely replicated results from our first study: lower AI knowledge was associated with higher risk scores while being younger and having higher confidence

(marginally significant) in AI knowledge was associated with higher benefit scores. Furthermore, results did not differ for both scores (RB1 and RB2) when controlling for participants' attitudes towards AI: Regarding the first measurement, RB1 (benefits outweigh the risks), age and knowledge were again significant predictors (note however, that effect sizes for knowledge and age were negligible). Additionally, and unsurprisingly – positive attitudes toward AI applications were associated with a higher RB1 score. For the second measurement, RB2 (difference between the ratings) confidence remained a significant predictor even when attitudes were included, while there was no influence of attitudes on this measurement. When investigating risks and benefits separately, we replicated our results from Study 1 for the risk measurement: having a lower AI knowledge score and more negative attitudes towards AI was associated with higher risk scores. For the benefit measurement, lower education and having a more positive attitude towards AI were associated with higher benefit scores, with confidence no longer significant.

Exploratory analysis

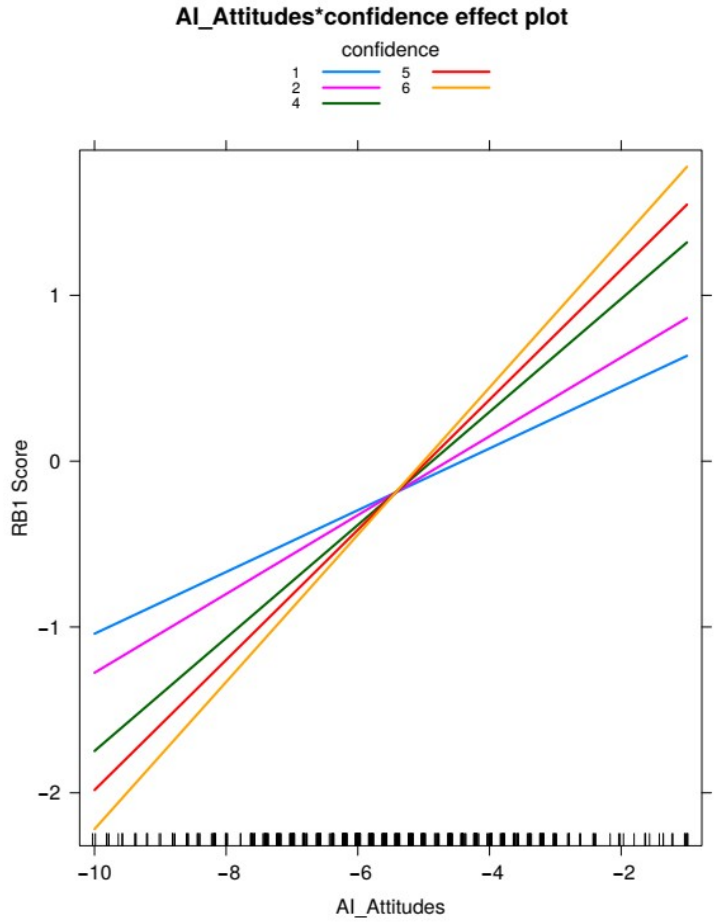
Investigating how confidence and attitudes towards AI influence people's perception of AI applications

In our research, we assumed that high confidence influences people's risk-benefit perception in a bi-directional manner, either biasing them towards a higher risk or a higher benefit assessment. And indeed, we found this result in both studies: higher (unwarranted) confidence led to a larger difference in people's risk-benefit perception. To gain a deeper understanding of what might drive this result, we conducted an exploratory analysis adding AI attitudes as a moderator variable. Given the importance of AI attitudes concerning perceptions of

AI (Araujo et al., 2020; Zhang & Dafoe, 2019), investigating the moderating role of AI attitudes thus, is a reasonable next step. We suspected that people with negative attitudes towards AI might rate the scenarios more as a risk when being (over)confident about their knowledge and vice versa: having positive attitudes towards AI and having high confidence regarding their knowledge might make people more likely to rate scenarios as benefits. To this end, we conducted a multiple regression for the RB1 score, adding confidence and attitude as an interaction term (variables were centered) while controlling for education, age, gender, and knowledge. The results indicate that a significant amount of variance in the RB1 score was explained by the predictors, $F(7, 428) = 18.09$, $p < .001$, $R^2 = .2$, $R^2_{Adjusted} = .2$. The interaction term was significant with b = .05, $t(428) = 2.13$, $p = .03$, Cohen's $f^2 = .01$, 95% CI [.00, .04]. Furthermore, age, b = -.01, $t(428) = -2.82$, $p = .005$, Cohen's $f^2 = .00$, 95% CI [.00, .03], gender, b =-.28, $t(428) = -2.28$, $p = .023$, Cohen's $f^2 = .03$, 95% CI [.01, .07], and knowledge, b = 1.19, $t(428) = 2.09$, $p = .037$, Cohen's $f^2 = .00$, 95% CI [.00, .04], significantly predicted the RB1 score. Thus, depending on the participant's attitudes towards AI, high confidence led to either benefits outweighing risks when attitudes were positive or risks outweighing benefits when attitudes were negative (see Figure 4, and Table E1, in the appendix).

Figure 4

The influence of confidence on the risk-benefit perception score RB1



Note. The figure displays the influence of confidence on the risk-benefit perception score RB1 with attitudes towards AI as moderator variable and age, gender, and education as covariates. Confidence levels are displayed in the top middle of the graph. For the interaction term variables were centered, however, to ease interpretation, the plot was generated based on the raw (non-centered) scores.

Investigating the role of metacognition regarding people's perception of AI applications

While some research has already shown that risk assessment is related to knowledge about the risk (Sjöberg & Drottz-Sjöberg, 1991; Gutteling & Kuttschreuter, 2011), and there is some evidence that confidence plays a role regarding risk assessment as well (Zhu et al., 2016), there have been no studies investigating the impact of participants' metacognitive abilities. Generally, the term metacognition describes insight into one's cognitive processes. More specifically, metacognition refers to people's insight, for example, into their knowledge, decision-making processes, or experiences (Flavell, 1979). Metacognition is thought to be linked to other concepts like consciousness (Koriat, 2007) and perceptual awareness (Fleming & Lau, 2014). However, there is also research questioning this proposed link between metacognition and perceptual awareness (Jachs et al., 2015).

Studies have yet to investigate the role of participants' metacognitive abilities, metacognitive sensitivity, and efficiency concerning their risk and benefit assessment. This lack of research is highly surprising, as metacognitive abilities play an essential role in a wide range of areas like decision-making (Allgood & Walstad, 2016; Balsdon et al., 2020; Desender et al., 2018; Hadar et al., 2013) learning (Veenman et al.,2006), and belief updating (Said et al., 2022; Fischer & Said, 2021). Thus, one could expect that accurately recognizing one's knowledge level about AI might also influence one's risk-benefit perception.

To this end, we asked the question if people's insight into the accuracy of their AI knowledge has an impact on how they perceive the AI scenarios: as a risk or as a benefit?

Leveraging methods from signal detection theory, we calculated metacognitive sensitivity as people's insight into the accuracy of their AI knowledge using the Bayesian approach from Fleming (2017). We also calculated group-level metacognitive efficiency: we looked at people's insight into the accuracy of their AI knowledge while controlling their performance in the AI knowledge test by applying a hierarchical Bayesian approach (Fleming, 2017). The participants' metacognitive efficiency was, on average, about 50% (M_{ratio_Study1} = .42 , CI_{Study1} = [.35, 0.49][5]; M_{ratio_Study2} = 0.61 ,CI_{Study2} = [0.52 , 0.73]). Average meta-d' was *M* = .41 (*SD* = 0.48) for Study 1 and *M* =.35 (*SD* = 0.65) for Study 2, which was rather low compared to their actual performance (especially in Study 1) on the AI knowledge test: with knowledge sensitivity d' being *M* = 0.9 (*SD* = 0.50) in Study 1 and *M* = 0.55 (*SD* = 0.66) in Study 2. Taken together, these results suggest that participants' insight into the accuracy of their AI knowledge was rather low.

In a next step, we investigated whether there was a relationship between metacognitive sensitivity and the risk-benefit measures: RB1, RB2, risk and benefits, adding metacognitive sensitivity as a predictor while controlling for age, gender, and education (as well as attitudes for Study 2). We only found a statistically significant relationship between metacognitive sensitivity and the RB2 score. For all other measures, we did not find any relationship. For the RB2 score, results for Study 1 showed that a significant amount of variance was explained by the investigated predictors, $F(4, 383) = 4.27$, $p = .002$, $R^2 = .04$, $R^2_{Adjusted} = .03$. The results also indicated that meta-d', b = .30, $t(383) = 3.30$, $p = .0011$, Cohen's $f^2 = .03$, 95% CI [.00, .07], was a significant predictor for the RB2 score (see Table D1, in the appendix). Results were

5 Note that CI here refers to the term "credible interval".

identical for Study 2, that is a significant amount of variance was explained by the investigated predictors, $F(5, 430) = 5.59$, $p < .001$, $R^2 = .06$, $R^2_{Adjusted} = .05$, with age, b =.01, $t(430) = 3.40$, $p < .001$, Cohen's $f^2 = .03$, 95% CI [.01, .07], and meta-d', b = .26, $t(430) = 3.57$, $p < .001$, Cohen's $f^2 = .03$, 95% CI [.01, .07], being significant predictors for the RB2 score (see Table D2, in the appendix).

Taken together, these results suggest that higher insight into the accuracy of the participants' AI knowledge leads to a bigger difference between risk-benefit scores. This result was rather surprising. We had expected that insight into the limitations of one's own knowledge about AI would lead people to rate scenarios in a more balanced way. This was not the case, however. Instead, higher insight leads participants, like higher confidence, to a less balanced assessment. One possible explanation for this finding could be that participants with low knowledge and high insight into their lack of knowledge ("knowing that I don't know") tend to overestimate the risks, executing caution, while participants with high knowledge and high insight into their AI knowledge accuracy ("knowing that I know") might overestimate the benefits, possibly contributing to *benefit amplification*.

Since individuals with high meta-d' values can have either (i) an insight into their AI knowledge or (ii) an insight into their lack of AI knowledge, without a clear distinction between the two, it could be a possible explanation for the lack of relationship with the other measurements. Attempting to gain more information, we ran an additional analysis investigating benefits and risks separately, with metacognitive sensitivity and knowledge sensitivity (both centered) as interaction terms, controlling for age, gender, education, and confidence. While the

interaction term was not significant for the risk score in both studies (with knowledge sensitivity again being the most important predictor for the risk assessment), the interaction term was significant for the benefit score in Study 2[6]. Participants with high knowledge and high insight into their AI knowledge accuracy ("knowing that I know") had a higher benefit score, with $F(7, 428) = 5.22$, $p < .001$, $R^2 = .08$, $R^2_{Adjusted} = .06$ and b = .17, $t(428) = 2.26$, $p = .025$, Cohen's $f^2 = .01$, 95% CI [.00, .04] for the interaction term. Furthermore, age, b = -.01, $t(428) = -3.95$, $p < .001$, Cohen's $f^2 = .03$, 95% CI [.01, .08], gender, b = -.18, t (428) = -2.13, $p = .03$, Cohen's $f^2 = .02$, 95% CI [.00, .05], and confidence b = .08, t (428) = 1.97, $p = .05$, Cohen's $f^2 = .01$, 95% CI [.00, .04], were significant predictors, with younger participants and male participants rating the benefits of the AI scenarios higher (see Table D3, in the appendix). Taken together, these results suggest that it is not only knowledge and confidence that influence people's risk-benefit perception when making judgements about AI applications, but also their insight into the accuracy of their AI knowledge.

General Discussion

Artificial intelligence applications are becoming increasingly ubiquitous in our everyday lives, playing a role in many areas, from driving to medicine. This makes investigating how AI applications are perceived by the general public of utmost importance. To this end, we conducted two studies ($N_{total} = 831$), with the second one being representative of the German population (age and gender), to gain insight into people's risk and benefit perception of AI applications in

6 Note that the interaction term remained significant when also controlling for attitudes. However, ΔR^2 was not significant anymore.

the areas of driving, media, and medicine. Employing a newly developed AI knowledge test as well as using a newly developed and validated benefit scale, we showed that AI knowledge and confidence in AI knowledge are predictors for people's risk-benefit perception of AI applications.

Influence of cognitive factors on risk and benefit perception of AI applications

Our study aimed to understand better how people's AI knowledge and their confidence in their AI knowledge are related to this ambiguity of AI. To this end, we presented participants with balanced scenarios (in terms of their associated risks and benefits) that allowed us to address, for the first time, how AI knowledge and confidence in AI knowledge are related to people's risk-benefit perception.

Participants generally performed similarly in both knowledge tests, with about 60% to 65% correct answers. That aligns with previous research showing that the public holds relatively little knowledge about AI and AI devices (Selwyn et al., 2020; Merenkov et al., 2021). At the same time, confidence was, on average, about 3 on a 5-point Likert scale. That is, participants responded on average that they are "70% sure" about their answers, thus, being slightly overconfident regarding their performance.

Regarding our AI use scenarios, we attempted to make them as balanced as possible regarding the associated risks and benefits. Some differences can still be seen, however, with benefits outweighing the risks in Study 1 and risks outweighing the benefits in Study 2. Despite these differences, the scenarios were largely rated as balanced: on a 5-point scale, the average risk perception and the average benefit perception were about 2.5 for both studies.

Is more knowledge always better? The case of risk blindness

So far, research on objective knowledge as a predictive factor for one's risk assessment is still sparse (Sjöberg & Drottz-Sjöberg, 1991), with some research indicating that low knowledge leads to judging risks as more severe. Furthermore, some recent studies showed a positive relationship between knowledge and benefit perception (Khare, 2023, topic-specific knowledge; Peng, 2020, general scientific knowledge). In our research we presented participants with balanced AI scenarios regarding the associated risks and benefits. That allowed us to investigate the relationship between knowledge and risk-benefit perception and to answer whether higher knowledge is related to a more *accurate* estimation of risks and benefits. While in the first study, only younger people perceived the scenarios as having higher associated benefits than risks, we found the proposed relationship between knowledge and risk-benefit assessment in the second study. For people with higher AI knowledge, benefits outweighed the risks.

To further understand what drives this result, an overestimation of the benefits, an underestimation of the risks, or both, we ran the same analysis separately for the risk and benefit scores. In both studies, higher knowledge led to a lower risk assessment, hinting to *risk blindness*. Furthermore, the results hold beyond participants' attitudes toward AI (Study 2). Thus, results suggest that higher levels of knowledge mainly affect the risk assessment, with risks possibly being underestimated. That is in line with research from Sjöberg & Drottz-Sjöberg (1991) in which lower knowledge was associated with higher risk assessment regarding nuclear power plants as well as literature on the differences between risk perception of experts and laypeople (Sjöberg, 1998; Sjöberg, 2002; Hansen et al., 2003; Sjöberg, 2008). Our research,

however, goes one step further. Since participants could assess both risks *and* benefits related to the scenarios presented, we see that higher knowledge only affected the risk assessment while having no effect on the benefit assessment.

As mentioned before, our study focused purely on the influence of knowledge and not expertise in AI. For future research, however, it might be interesting to see whether our results can be replicated when AI experts judge the scenarios. Additionally, a currently still open question is why higher levels of knowledge lead to risk blindness. One plausible explanation might be that people overestimate their abilities to mitigate the risks, thus rating them lower.

Metacognition and risk-benefit perception

Even though metacognition plays a significant role in a wide range of areas (Balsdon et al., 2020; Veenman et al.,2006; Said et al., 2022), there is still a lack of research on how people's metacognitive abilities relate to their risk-benefit assessment. In our research, we investigated this relationship with respect to the three facets of metacognition: metacognitive bias (confidence), metacognitive sensitivity, and metacognitive efficiency.

Metacognitive bias (confidence)

While there is some literature on the subjective ratings of knowledge and risk perception (Jaccard et al., 2005; Zhu, 2016), there is, to our knowledge, little or no research assessing confidence in AI knowledge concerning risk-benefit perception. We addressed this research gap by showing that higher (unwarranted) confidence leads to a larger difference in people's risk-benefit perception. That is, having higher confidence in one's knowledge leads people to either overestimate benefits or overestimate risks. To gain further insight into what might be driving

these findings, we included participants' attitudes toward AI as a moderator variable. Results show an interaction between participants' attitudes toward AI and confidence on the RB1 score (RB1 was calculated as the difference between the benefit and the risk score). Having positive attitudes towards AI and high confidence in one's knowledge leads participants to see benefits as outweighing the risks. The risks outweighed the benefits for participants who reported more negative attitudes toward AI and high confidence in their knowledge. Taken together, we showed that (i) confidence in AI knowledge plays a significant role in people's risk-benefit perception and (ii) that attitudes towards AI are at least one factor that explains the bi-directional nature of high confidence concerning risk-benefit perception.

When investigating risks and benefits separately, we found that higher confidence in AI knowledge was associated with higher benefit scores. However, this finding was only found in Study 1 and could not be replicated in Study 2. For Study 2, results were only marginally significant and not significant when controlling for attitudes toward AI. Thus, it seems confidence plays an important role, especially regarding the *difference between the ratings*: higher confidence leads to a less balanced risk-benefit assessment.

Metacognitive sensitivity

Over-confidence regarding the accuracy of one's knowledge can have dire consequences; some recent research has pointed toward a relationship between knowing the limitations of one's knowledge and more accurate decision-making (Fischer et al., 2021). Thus, one expectation was that insight into the limitations of one's knowledge about AI would lead people to rate scenarios

in a more balanced way. However, this was not the case: instead, higher insight into the limitations of one's knowledge led – like higher confidence – to a less balanced assessment.

On the one hand, we assumed that participants with low knowledge and high insight into their lack of knowledge ("knowing that I don't know") would tend to overestimate the risks making them more cautious. On the other hand, we expected that participants with high knowledge and high insight into their AI knowledge accuracy ("knowing that I know") would tend to overestimate the benefits (benefit amplification). While we did not find the proposed relationship for the risk assessment, results showed, at least for Study 2, that participants with high knowledge and high insight into their AI knowledge accuracy ("knowing that I know") had a higher benefit score. That is highly interesting, as for the knowledge score, the opposite was true: higher knowledge leads people to underestimate the risks.

Thus, it seems like cognitive abilities (knowledge) affect only the risk score, resulting in risk blindness, while metacognitive abilities (confidence) affect only the benefit score leading to benefit amplification. However, as for the latter, the results were only significant for Study 2. Thus, further research to corroborate this result is needed.

Metacognitive efficiency

Metacognitive efficiency is a more general measure of people's insight into the accuracy of their knowledge as it controls for task performance. Thus, to understand people's metacognitive abilities, we used the hierarchical Bayesian procedure of Fleming (2017) to calculate Mratio for both studies. Results showed that participants' metacognitive efficiency was, on average, about 50%. Interestingly, when comparing these findings to results from other

knowledge tests, Mratio for AI knowledge was similar to Mratio for climate change knowledge (~50%, Fischer et al., 2019), while for science knowledge, Mratio was about ~100% and ~86 % for COVID-19 knowledge (Fischer et al., 2021). So far, to our knowledge, there is no research explaining the differences in metacognitive efficiency for the different knowledge tests. Fleming (2021) argued that some areas of knowledge might be especially susceptible to what he terms metacognitive blind spots. However, why and under what circumstances these blind spots might occur is still an open question.

Risk and benefits: two sides of the same coin?

While risks and benefits are negatively correlated in literature (Alhakami & Slovic, 1994; Fischhoff et al., 1979; Slovic et al., 1991), the ambiguous nature of AI applications could very well conflict with the current consensus, resulting in distinct mental concepts of what AI risks and AI benefits are. Usually, perceived AI risks center around the misuse of AI by the government or tech companies, while perceived AI benefits might be more related to the application's content. In our research, risks, and benefits were negatively correlated in both studies. However, correlation coefficients were relatively low, with r = -.42 (Study1) and r = -.18 (Study 2). While one could argue that it is not necessary to assess risks and benefits separately, our findings showed that it is vital – at least for the topic of AI – to assess both, as knowledge and confidence were differently related to the risk and benefit measures. To better understand whether AI-related risks and benefits are perceived as the same or a different construct, future studies should address this question.

Limitations

There are two limitations regarding our research. First, even though we did improve the alpha in our AI knowledge test in the second study, item quality still needs to be enhanced, ideally aiming at an alpha value of 0.8. Second, participants only received one scenario each. As discussed earlier, especially in case of the media scenarios, there might be a content dependency regarding the risk-benefit assessment. Thus, future studies should show multiple scenarios for each participant to make a more general statement about people's risk-benefit perceptions.

Conclusion

Our research investigated the influence of people's AI knowledge and confidence in their AI knowledge on AI risk-benefit perception. We showed that higher confidence in AI knowledge and attitudes towards AI bias people to either overestimate the risks or the benefits, showing a possible bi-directional influence. We showed that people exhibit *risk blindness* for higher levels of AI knowledge and, equally important, that people with high confidence exhibit *benefit amplification.* In a society where AI applications play an undeniably increasingly central role, our research is an essential first step towards understanding more about people's perceptions of said AI applications and what cognitive factors might drive and shape these perceptions. Since knowledge and confidence affect risk and benefit, respectively, one must wonder if the phenomena of *risk blindness* and *benefit amplification* carry over to not only a cross-cultural context but to AI experts as well. Our research highlights a need to improve the understanding among the general population of AI-related risk and benefit perception as well as, ideally, a range of other groups who play an active role in AI policy-making, such as politicians and AI experts, with the imminent ability to either advance or deter the spread of AI technology.

References

Ahmad, M. (2021). Does underconfidence matter in short-term and long-term investment decisions? Evidence from an emerging market. *Management Decision*, *59*(3), 692–709. https://doi.org/10.1108/MD-07-2019-0972

Alhakami, A. S., & Slovic, P. (1994). A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis*, *14*(6), 1085–1096. https://doi.org/10.1111/j.1539-6924.1994.tb00080.x

Allgood, S., & Walstad, W. B. (2016). The effects of perceived and actual financial literacy on financial behaviors: Perceived and financial literacy. *Economic Inquiry*, *54*(1), 675–697. https://doi.org/10.1111/ecin.12255

Antes, A. L., Burrous, S., Sisk, B. A., Schuelke, M. J., Keune, J. D., & DuBois, J. M. (2021). Exploring perceptions of healthcare technologies enabled by artificial intelligence: an online, scenario-based survey. *BMC Medical Informatics and Decision Making*, 21(1), 1-15. https://doi.org/10.1186/s12911-021-01586-8

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, *11*(1), 1753. https://doi.org/10.1038/s41467-020-15561-w

Bao, L., Krause, N. M., Calice, M. N., Scheufele, D. A., Wirz, C. D., Brossard, D., ... & Xenos, M. A. (2022). Whose AI? How different publics think about AI and its social impacts. Computers in Human Behavior, 130, 107182. https://doi.org/10.1016/j.chb.2022.107182.

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

Bartlett, R. (2006). Artificial intelligence in sports biomechanics: New dawn or false hope? *Journal of Sports Science & Medicine*, *5*(4), 474–479.

Beets, B., Newman, T. P., Howell, E. L., Bao, L., & Yang, S. (2023). Surveying Public Perceptions of Artificial Intelligence in Health Care in the United States: Systematic Review. *Journal of Medical Internet Research*, 25, e40337. https://doi.org/10.2196/40337

Bertoldo, R., Mays, C., Poumadère, M., Schneider, N., & Svendsen, C. (2016). Great deeds or great risks? Scientists' social representations of nanotechnology. *Journal of Risk Research*, *19*(6), 760–779. https://doi.org/10.1080/13669877.2015.1042503

Brossard, D., & Nisbet, M. C. (2007). Deference to scientific authority among a low information public: Understanding US opinion on agricultural biotechnology. *International Journal of Public Opinion Research*, *19*(1), 24-52. https://doi.org/10.1093/ijpor/edl003

Broihanne, M. H., Merli, M., & Roger, P. (2014). Overconfidence, risk perception and the risk-taking behavior of finance professionals. *Finance Research Letters*, *11*(2), 64–73. https://doi.org/10.1016/j.frl.2013.11.002

Callaghan, V., Clarke, G., Colley, M., & Hagras, H. (2001). A soft-computing distributed artificial intelligence architecture for intelligent buildings. In J. Kacprzyk, V. Loia, & S. Sessa (Eds.), *Soft Computing Agents* (Vol. 75, pp. 117–145). Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-1815-4_5

Canfield, C., Fischhoff, B. & Davis, A. Better beware: comparing metacognition for phishing and legitimate emails. *Metacognition Learning* 14, 343–362 (2019). https://doi.org/10.1007/s11409-019-09197-5

Cassam, Q. (2017). Diagnostic error, overconfidence and self-knowledge. *Palgrave Communications*, *3*(1), 17025. https://doi.org/10.1057/palcomms.2017.25

Chen, Y., & Li, Y. (2019). Intelligent autonomous pollination for future farming—A micro air vehicle conceptual framework with artificial intelligence and human-in-the-loop. *IEEE Access*, *7*, 119706–119717. https://doi.org/10.1109/ACCESS.2019.2937171

Cochran, W. G. (1977). *Sampling techniques* (3d ed). Wiley.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Cui, D., & Wu, F. (2021). The influence of media use on public perceptions of artificial intelligence in China: evidence from an online survey. *Information Development*, 37(1), 45-57. https://doi.org/10.1177/02666669198934

de Almeida, A. F., Moreira, R., & Rodrigues, T. (2019). Synthetic organic chemistry driven by artificial intelligence. *Nature Reviews Chemistry*, *3*(10), 589–604. https://doi.org/10.1038/s41570-019-0124-0

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, *40*, 434-449. https://doi.org/10.1007/s11747-011-0300-3

Dunjko, V., & Briegel, H. J. (2018). Machine learning & artificial intelligence in the quantum domain: A review of recent progress. *Reports on Progress in Physics*, *81*(7), 074001. https://doi.org/10.1088/1361-6633/aab406

Edelman AI Survey. *Survey of technology executives and the general population shows excitement and curiosity yet uncertainty and worries that artificial intelligence could be a tool of division*. (2019). Retrieved March 3rd, 2022 from https://www.edelman.com/research/2019-artificial-intelligence-survey

Fast, E., & Horvitz, E. (2016). Long-term trends in the public perception of artificial intelligence. *ArXiv:1609.04904 [Cs]*. http://arxiv.org/abs/1609.04904

Fischer, H., Amelung, D., & Said, N. (2019). The accuracy of German citizens' confidence in their climate change knowledge. *Nature Climate Change*, *9*(10), 776-780. https://doi.org/10.1038/s41558-019-0563-0

Fischer, H., & Said, N. (2021). Importance of domain-specific metacognition for explaining beliefs about politicized science: The case of climate change. *Cognition, 208,* 104545. https://doi.org/10.1016/j.cognition.2020.104545

Fischer, H., Huff, M., & Said, N. (2021, March 21). Insight into the accuracy of COVID-19 beliefs predicts behavior during the pandemic. https://doi.org/10.31234/osf.io/x2qv3

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979). Weighing the risks: Risks: Benefits which risks are acceptable? *Environment: Science and Policy for Sustainable Development*, *21*(4), 17–38. https://doi.org/10.1080/00139157.1979.9929722

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist,* 34(10), 906. https://doi.org/10.1037/0003-066X.34.10.906

Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1). https://doi.org/10.1093/nc/nix007

Fleming, S. M. (2021). *Know thyself: How the new science of self-awareness*. John Murray.

Gartner Survey Shows 37 Percent of Organizations Have Implemented #AI. (2019, January 21). Gartner. Retrieved on March 3rd ,2022 from https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have.

Gerd Gigerenzer, Jutta Mata, Ronald Frank, Public Knowledge of Benefits of Breast and Prostate Cancer Screening in Europe, *JNCI: Journal of the National Cancer Institute,* Volume 101, Issue 17, 2 September 2009, Pages 1216–1220, https://doi.org/10.1093/jnci/djp237

Gershoff, A. D., & Johar, G. V. (2006). Do you know me? Consumer calibration of friends' knowledge. *Journal of Consumer Research*, *32*(4), 496–503. https://doi.org/10.1086/500479

Glaser, M., & Weber, M. (2007). Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, *32*(1), 1–36. https://doi.org/10.1007/s10713-007-0003-3

Goldsmith, R. E., & Pillai, K. G. (2008). Knowledge calibration. In M. E. Jennex (Ed.), *Knowledge Management* (pp. 3125–3132). IGI Global. https://doi.org/10.4018/978-1-59904-933-5.ch256

Goto-Tamura, R., Takesue, Y., & Takesue, S. (1976). Immunological similarity between NADH-cytochrome b5 reductase of erythrocytes and liver microsomes. *Biochimica Et Biophysica Acta*, *423*(2), 293–302. https://doi.org/10.1016/0005-2728(76)90186-9

Gutteling, J. M., & Kuttschreuter, M. (2002). The role of expertise in risk communication: Laypeople's and expert's perception of the millennium bug risk in The Netherlands. *Journal of Risk Research*, *5*(1), 35–47. https://doi.org/10.1080/13669870010029639

Hadar, L., Sood, S., & Fox, C. R. (2013). Subjective knowledge in consumer financial decisions. *Journal of Marketing Research*, *50*(3), 303–316. https://doi.org/10.1509/jmr.10.0518

Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, S36-S40. https://doi.org/10.1016/j.metabol.2017.01.011

Hansen, J., Holm, L., Frewer, L., Robinson, P., & Sandøe, P. (2003). Beyond the knowledge deficit: Recent research into lay and expert attitudes to food risks. *Appetite*, *41*(2), 111–121. https://doi.org/10.1016/S0195-6663(03)00079-5

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide* (3rd ed.). Routledge. https://doi.org/10.4324/9781003203636

He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine,* 25(1), 30-36. https://doi.org/10.1038/s41591-018-0307-0

Iorfa, S. K., Ottu, I. F. A., Oguntayo, R., Ayandele, O., Kolawole, S. O., Gandi, J. C., Dangiwa, A. L., & Olapegba, P. O. (2020). COVID-19 knowledge, risk perception, and precautionary behavior among Nigerians: A moderated mediation approach. *Frontiers in Psychology*, *11*, 566773. https://doi.org/10.3389/fpsyg.2020.566773

Jaccard, J., Dodge, T., & Guilamo-Ramos, V. (2005). Metacognition, risk behavior, and risk outcomes: The role of perceived intelligence and perceived knowledge. *Health Psychology*, *24*(2), 161–170. https://doi.org/10.1037/0278-6133.24.2.161

Jachs, B., Blanco, M. J., Grantham-Hill, S., and Soto, D. (2015). On the independence of visual awareness and metacognition: a signal detection theoretic analysis. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 269–276. https://doi.org/10.1037/xhp0000026

Janowski, K., Ritschel, H., Lugrin, B., & André, E. (2018). *Sozial interagierende Roboter in der Pflege* (pp. 63-87). Springer Fachmedien Wiesbaden.

Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*, *2*, 1–12. https://doi.org/10.1016/j.aiia.2019.05.004

Jones, E. J., Appel, S. J., Eaves, Y. D., Moneyham, L., Oster, R. A., & Ovalle, F. (2012). Cardiometabolic risk, knowledge, risk perception, and self-efficacy among American Indian

women with previous gestational diabetes. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, *41*(2), 246–257. https://doi.org/10.1111/j.1552-6909.2012.01339.x

Khare, A. (2023). Green apparel buying: Role of past behavior, knowledge and peer influence in the assessment of green apparel perceived benefits. *Journal of International Consumer Marketing*, *35*(1), 109-125. https://doi.org/10.1080/08961530.2019.1635553

Kelley, P. G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., Kramm, A., Newman, D. T., & Woodruff, A. (2021). Exciting, useful, worrying, futuristic: Public perception of artificial intelligence in 8 countries. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 627–637. https://doi.org/10.1145/3461702.3462605

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 289–325). Cambridge University Press. https://doi.org/10.1017/CBO9780511816789.012

Kozyreva, A., Herzog, S., Lorenz-Spreen, P., Hertwig, R., & Lewandowsky, S. (2020). *Artificial intelligence in online environments: Representative survey of public attitudes in Germany* (p. 2150178). Max Planck Institute for Human Development. https://doi.org/10.17617/2.3188061

Kozyreva, A., Lorenz-Spreen, P., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the United States. *Humanities and Social Sciences Communications*, *8*(1), 117. https://doi.org/10.1057/s41599-021-00787-w

Kuttschreuter, M., Gutteling, J. M., & de Hond, M. (2011). Framing and tone-of-voice of disaster media coverage: The aftermath of the Enschede fireworks disaster in the Netherlands. *Health, Risk & Society*, *13*(3), 201–220. https://doi.org/10.1080/13698575.2011.558620

Larsson, S., Jansson, M., & Boholm, Å. (2019). Expert stakeholders' perception of nanotechnology: Risk, benefit, knowledge, and regulation. *Journal of Nanoparticle Research*, *21*(3), 57. https://doi.org/10.1007/s11051-019-4498-1

Li, Y., Wang, J., & Rao, H. R. (2022, February). Metacognitive Skills in Phishing Email Detection: A Study of Calibration and Resolution. In *Secure Knowledge Management In The Artificial Intelligence Era: 9th International Conference, SKM 2021, San Antonio, TX, USA, October 8–9, 2021, Proceedings* (pp. 37-47). Cham: Springer International Publishing.

Liang, Y., Lee, S.-H., & Workman, J. E. (2020). Implementation of artificial intelligence in fashion: Are consumers ready? *Clothing and Textiles Research Journal*, *38*(1), 3–18. https://doi.org/10.1177/0887302X19873437

Malmendier, U., & Tate, G. (2015). Behavioral CEOs: The role of managerial overconfidence. *Journal of Economic Perspectives*, *29*(4), 37–60. https://doi.org/10.1257/jep.29.4.37

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Mannes, A. E., & Moore, D. A. (2013). A behavioral demonstration of overconfidence in judgment. *Psychological Science*, *24*(7), 1190–1197. https://doi.org/10.1177/0956797612470700

Maswanya, E. S. (1999). Knowledge, risk perception of AIDS and reported sexual behaviour among students in secondary schools and colleges in Tanzania. *Health Education Research*, *14*(2), 185–196. https://doi.org/10.1093/her/14.2.185

McKenzie, C. R. M., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes*, *107*(2), 179–191. https://doi.org/10.1016/j.obhdp.2008.02.007

Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In V. C. Müller (Ed.), *Fundamental Issues of Artificial Intelligence* (Vol. 376, pp. 555–572). Springer International Publishing. https://doi.org/10.1007/978-3-319-26485-1_33

Murad, Z., Sefton, M., & Starmer, C. (2016). How do risk attitudes affect measured confidence? *Journal of Risk and Uncertainty*, *52*(1), 21–46. https://doi.org/10.1007/s11166-016-9231-1

Naguib, M., Brull, S. J., Hunter, J. M., Kopman, A. F., Fülesdi, B., Johnson, K. B., & Arkes, H. R. (2019). Anesthesiologists' overconfidence in their perceived knowledge of neuromuscular monitoring and its relevance to all aspects of medical practice: An international survey. *Anesthesia & Analgesia*, *128*(6), 1118–1126. https://doi.org/10.1213/ANE.0000000000003714

Nisbet, M. C., Scheufele, D. A., Shanahan, J., Moy, P., Brossard, D., & Lewenstein, B. V. (2002). Knowledge, reservations, or promise? A media effects model for public perceptions of science and technology. *Communication Research*, 29(5), 584-608. https://doi.org/10.1177/00936500223619

Nosić, A., & Weber, M. (2010). How riskily do I invest? The role of risk attitudes, risk perceptions, and overconfidence. *Decision Analysis*, *7*(3), 282–301. https://doi.org/10.1287/deca.1100.0178

Oliver, D. C. (2022). *Americans are concerned about rise of AI and human enhancements*. CNN. Retrieved March 30, 2022, from https://edition.cnn.com/2022/03/17/health/ai-human-enhancements/index.html

Park, Y. J., & Jones-Jang, S. M. (2022). Surveillance, security, and AI as technological acceptance. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01331-9

Park, J., Hong, E., & Le, H. T. (2021). Adopting autonomous vehicles: The moderating effects of demographic variables. *Journal of Retailing and Consumer Services*, *63*, 102687. https://doi.org/10.1016/j.jretconser.2021.102687

Peng, Y. (2020). The ideological divide in public perceptions of self-driving cars. *Public Understanding of Science*, *29*(4), 436-451. https://doi.org/10.1177/0963662520917339

Pengnate, S., & Antonenko, P. (2013). A Multimethod Evaluation of Online Trust and Its Interaction with Metacognitive Awareness: An Emotional Design Perspective. *International Journal of Human–Computer Interaction*, *29*(9), 582–593. https://doi.org/10.1080/10447318.2012.735185

Pikulina, E., Renneboog, L., & Tobler, P. N. (2017). Overconfidence and investment: An experimental approach. *Journal of Corporate Finance*, *43*, 175–192. https://doi.org/10.1016/j.jcorpfin.2017.01.002

Rahi, S., Alnaser, F. M., & Abd Ghani, M. (2019). Designing survey research: recommendation for questionnaire development, calculating sample size and selecting research paradigms. *Economic and Social Development: Book of Proceedings*, 1157-1169.

Rathi, S. K., D. S. ,. Rathi. (2019). Knowledge on artificial intelligence and related fields among engineering students. *The International Journal of Engineering and Science (IJES)*, *8*(2), 03–10. https://doi.org/DOI:10.9790/1813-0802010310

Rohden, S. F., & Zeferino, D. G. (2022). Recommendation agents: an analysis of consumers' risk perceptions toward artificial intelligence. *Electronic Commerce Research*, 1-16. https://doi.org/10.1007/s10660-022-09626-9

Robinson, A. T., & Marino, L. D. (2015). Overconfidence and risk perceptions: Do they really matter for venture creation decisions? *International Entrepreneurship and Management Journal*, *11*(1), 149–168. https://doi.org/10.1007/s11365-013-0277-0

Rundmo, T., & Moen, B. (2006). Risk perception and demand for risk mitigation in transport: A comparison of lay people, politicians and experts. *Journal of Risk Research*, *9*(6), 623–640. https://doi.org/10.1080/13669870600813811

Said, N., Fischer, H., & Anders, G. (2022). Contested science: Individuals with higher metacognitive insight into interpretation of evidence are less likely to polarize. *Psychonomic Bulletin & Review*, *29*(2), 668–680. https://doi.org/10.3758/s13423-021-01993-y

Sarno, D. M., & Neider, M. B. (2022). So many phish, so little time: Exploring email task factors and phishing susceptibility. *Human Factors*, 64(8), 1379-1403. https://doi.org/10.1177/0018720821999174

Sarno, D. M., McPherson, R., & Neider, M. B. (2022). Is the key to phishing training persistence?: Developing a novel persistent intervention. *Journal of Experimental Psychology: Applied, 28*(1), 85–99. https://doi.org/10.1037/xap0000410

Savadori, L., Savio, S., Nicotra, E., Rumiati, R., Finucane, M., & Slovic, P. (2004). Expert and public perception of risk from biotechnology. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *24*(5), 1289–1299. https://doi.org/10.1111/j.0272-4332.2004.00526.x

Schumm, H. (2021). Fragebogen zur Erhebung des Wissensstandes über Küstliche Intelligenz [Unpublished bachelor's thesis]. University of Tübingen.

Selwyn, N., & Gallo Cordoba, B. (2021). Australian public understandings of artificial intelligence. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01268-z

Selwyn, N., Cordoba, B. G., Andrejevic, M., & Campbell, L. (2020). *AI for social good—Australian attitudes toward AI and society report.pdf* (p. 7282677 Bytes). Monash University. https://doi.org/10.26180/13159781.V1

Siegrist, M., Keller, C., & Kiers, H. A. (2005). A new look at the psychometric paradigm of perception of hazards. *Risk Analysis: An International Journal*, *25*(1), 211-222. https://doi.org/10.1111/j.0272-4332.2005.00580.x

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI - Künstliche Intelligenz*, *35*(1), 109–118. https://doi.org/10.1007/s13218-020-00689-0

Sjöberg, L. (1998). Risk perception: Experts and the public. *European Psychologist*, *3*(1), 1–12. https://doi.org/10.1027//1016-9040.3.1.1

Sjöberg, L. (1998). Worry and risk perception. *Risk Analysis*, *18*(1), 85–93. https://doi.org/10.1111/j.1539-6924.1998.tb00918.x

Sjöberg, L. (2000). The methodology of risk perception research. *Quality and Quantity*, *34*(4), 407–418. https://doi.org/10.1023/A:1004838806793

Sjöberg, L. (2002). The allegedly simple structure of experts' risk perception: An urban legend in risk research. *Science, Technology, & Human Values*, *27*(4), 443–459. https://doi.org/10.1177/016224302236176

Sjöberg, L. (2008). Genetically modified food in the eyes of the public and experts. *Risk Management*, *10*(3), 168–193. https://doi.org/10.1057/rm.2008.2

Sjoberg, L., & Drottz-Sjoberg, B.-M. (1991). Knowledge and risk perception among nuclear power plant employees. *Risk Analysis*, *11*(4), 607–618. https://doi.org/10.1111/j.1539-6924.1991.tb00650.x

Sjöberg, L., & Wåhlberg, A. af. (2002). Risk perception and new age beliefs. *Risk Analysis*, *22*(4), 751–764. https://doi.org/10.1111/0272-4332.00066

Smith, A., & Anderson, J. (2019, December 31). *Predictions for the State of AI and Robotics in 2025*. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2014/08/06/predictions-for-the-state-of-ai-and-robotics-in-2025/

Statista. (2021, August 18). *Virtual Assistant Technology - Statistics & Facts*. Retrieved March 3, 2022, from https://www.statista.com/topics/5572/virtual-assistants/#topicHeader__wrapper

Storer, H. L., Scott, C. F., Rodriguez, M., & Nyerges, E. X. (2023). Technology Is a "Blessing and a Curse": The Perceived Risks and Benefits of Digital Technology Adoption at Domestic Violence

Organizations that Serve Teens. *Journal of Technology in Human Services*, 41(1), 96-124. https://doi.org/10.1080/15228835.2023.2179158

Sütterlin, S., Lugo, R., Ask, T., Veng, K., Eck, J., Fritschi, J., Talha-Özmen, M., Bärreiter, B., & Knox, B. (2022). The Role of IT Background for Metacognitive Accuracy, Confidence and Overestimation of Deep Fake Recognition Skills. *Lecture Notes in Computer Science*, *13310*, 103–119. https://doi.org/10.1007/978-3-031-05457-0_9

Tussyadiah, I., & Miller, G. (2019). Perceived impacts of artificial intelligence and responses to positive behaviour change intervention. In *Information and Communication Technologies in Tourism 2019: Proceedings of the International Conference* in Nicosia, Cyprus, January 30–February 1, 2019 (pp. 359-370). Springer International Publishing.

Veenman, M. V. J. (2007). The assessment and instruction of self-regulation in computer-based environments: A discussion. *Metacognition and Learning*, *2*(2–3), 177–183. https://doi.org/10.1007/s11409-007-9017-6

Walpole, H. D., & Wilson, R. S. (2021). Extending a broadly applicable measure of risk perception: The case for susceptibility. *Journal of Risk Research*, *24*(2), 135–147. https://doi.org/10.1080/13669877.2020.1749874

Wilson, R. S., Zwickle, A., & Walpole, H. (2019). Developing a broadly applicable measure of risk perception. *Risk Analysis*, 39(4), 777–791. https://doi.org/10.1111/risa.13207

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*, 292(1), 60-66. https://doi.org/10.1148/radiol.2019182716

Yang, H., & Thompson, C. (2010). Nurses' risk assessment judgements: A confidence calibration study: Nurses' risk assessment judgements. *Journal of Advanced Nursing*, *66*(12), 2751–2760. https://doi.org/10.1111/j.1365-2648.2010.05437.x

Yang, H., Thompson, C., & Bland, M. (2012). Effect of improving the realism of simulated clinical judgement tasks on nurses' overconfidence and underconfidence: Evidence from a comparative confidence calibration analysis. *International Journal of Nursing Studies*, *49*(12), 1505–1511. https://doi.org/10.1016/j.ijnurstu.2012.08.005

Yeh, S. C., Wu, A. W., Yu, H. C., Wu, H. C., Kuo, Y. P., & Chen, P. X. (2021). Public perception of artificial intelligence and its connections to the sustainable development goals. *Sustainability*, 13(16), 9165. https://doi.org/10.3390/su13169165

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, B., & Dafoe, A. (2019). Artificial intelligence: American attitudes and trends. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3312874

Zhang, B., & Dafoe, A. (2020). U.S. public opinion on the governance of artificial intelligence. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 187–193. https://doi.org/10.1145/3375627.3375827

Zhang, J., While, A. E., & Norman, I. J. (2012). Seasonal influenza vaccination knowledge, risk perception, health beliefs and vaccination behaviours of nurses. *Epidemiology and Infection*, *140*(9), 1569–1577. https://doi.org/10.1017/S0950268811002214

Zhu, W., Wei, J., & Zhao, D. (2016). Anti-nuclear behavioral intentions: The role of perceived knowledge, information processing, and risk perception. *Energy Policy*, *88*, 168–177. https://doi.org/10.1016/j.enpol.2015.10.009

Declarations

Ethics approval and consent to participate

APA ethical standards were followed in the conduct of both studies reported in this article and informed consent was collected from the participants at the beginning of the study. Both studies were approved by the ethics committee of the Leibniz-Institut für Wissensmedien, Tübingen.

Availability of data and material

Study1. All hypotheses were pre-registered under https://aspredicted.org/CFS_HBW. Additional analyses are indicated as exploratory. All data, materials, and the analysis code (in R) are publicly available under: https://osf.io/mbg79/. All participants provided informed consent. Methods were approved by the ethical review board of the Leibniz-Institut für Wissensmedien in Tübingen, Germany.

Study2. All hypotheses were pre-registered under https://aspredicted.org/34B_R2F. Additional analyses are indicated as exploratory. All data, materials, and analysis code (in R) are publicly available under: https://osf.io/mbg79/. All participants provided informed consent. Methods were approved by the ethical review board of the Leibniz-Institut für Wissensmedien in Tübingen, Germany.

Competing interests

The authors declare that they have no competing interests.

Funding

Data collection was funded by internal funds of the Leibniz-Institut für Wissensmedien

(Tübingen, Germany).

Acknowledgments: We thank Nicole Antes for supporting the development of the risk-benefit scenarios.

Authors' contributions

Conceptualization: NS, IB, JB, MH

Study implementation: HS, AP, NS

Formal analysis: NS, HS

Visualization: NS

Writing – original draft: NS, AP

Writing – review & editing: NS, AP, IB, JB, MH

Appendix A

Table A1

Items used to measure knowledge level on Artificial Intelligence

Domains	Number	Original German Items	English Translated Items
Functioning	1a***	Künstliche Intelligenz bezeichnet Computer, die selbständig Entscheidungen treffen können und lernfähig sind. (Richtig)	Artificial intelligence refers to computers that can make decisions independently and are capable of learning. (True)
	2a***	Ein neuronales Netzwerk kann sich durch Lernen und Training verbessern. (Richtig)	A neural network can improve through learning and training. (True)
	3a***	Der Turing-Test wird dazu verwendet, festzustellen, ob eine Künstliche Intelligenz als intelligent wahrgenommen wird. (Richtig)	The Turing test is used to determine whether an artificial intelligence is perceived as intelligent. (True)
	4a***	Eine starke KI kann selbstständig Entscheidungen treffen. (Richtig)	A strong AI can make decisions on its own. (True)
	5a***	Eine schwache KI kann Profi-Schachspieler schlagen. (Richtig)	A weak AI can beat professional chess players. (True)
	6a*	Die Künstlichen Intelligenzen, die in den letzten Jahren zum Einsatz kommen, werden als starke KIs bezeichnet. (Falsch)	The artificial intelligences that have been deployed in recent years are referred to as strong AIs. (False)
	7a***	Maschinelles Lernen basiert auf statistischen Prinzipien. (Richtig)	Machine learning is based on statistical principles. (True)
	8a***	Beim Maschinellen Lernen können zwei Arten an Algorithmen unterschieden werden: überwachtes und unüberwachtes Lernen. (Richtig)	In machine learning, two types of algorithms can be distinguished: supervised and unsupervised learning. (True)
	9a*	Deep Learning braucht geringere Datenmengen als andere Formen der KI. (Falsch)	Deep Learning requires smaller amounts of data than other forms of AI. (False)
	10a*	Bei künstlichen neuralen Netzwerken wird versucht, die neuralen Netzwerke im Gehirn vollständig nachzubilden. (Falsch)	Artificial neural networks attempt to fully replicate neural networks in the brain. (False)

	11a***	Autonome militärische Drohnen, die ihre Ziele selbst bestimmen und angreifen, werden in kriegerischen Konflikten eingesetzt. (Richtig)	Autonomous military drones that determine and attack targets on their own are used in warlike conflicts. (True)
	1b**	KI-Systeme schneiden bei der richtigen Erkennung von Bildern schlechter ab als Menschen. (Falsch)	AI systems perform worse than humans at correctly recognizing images. (False)
	2b**	Es gibt noch keine KI-basierten Roboter, die nur durch visuelle Demonstration lernen (z.B. einen Turm zu bauen). (Falsch)	There are no AI-based robots yet, that learn only through visual demonstration (e.g., to build a tower). (False)
	3b**	YOLO bedeutet im KI-Kontext “You Only Learn Once”. (Falsch)	YOLO in the AI context means "You Only Learn Once." (False)
	4b**	Um eine Aufgabe zu lernen, benötigen KI-Systeme nur sehr wenige Daten, um an die Leistung von Menschen heranzukommen. (Falsch)	To learn a task, AI systems need very little data to reach the level of human performance.(False)
	5b**	Maschinelles Lernen ist ein Überbegriff für künstliche Intelligenz. (Falsch)	Machine learning is an umbrella term for artificial intelligence. (False)
	6b**	CNN – Als Convolutional Neural Networks werden neuronale Netzwerke bezeichnet, die Bilder verändern, z.B. sie neu einfärben. (Falsch)	CNN - Convolutional neural networks are neural networks that alter images, such as recoloring them. (False)
	7b**	Ein vereinfachtes künstliches neuronales Netz wird auch Nervenbahn genannt. (Falsch)	A simplified artificial neural network is also called a neural pathway. (False)
Medical Field	12a***	Mittels Künstlicher Intelligenz können Krebsdiagnosen früher gestellt werden. (Richtig)	Artificial intelligence can be used to diagnose early cancer. (True)
	13a	KI kann keine individuellen Therapievorschläge entwerfen. (Falsch)	AI cannot design individualized therapy suggestions. (False)
	14a***	Krankenhäuser/Ärzte arbeiten bereits mit KI, um Diagnosen zu treffen. (Richtig)	Hospitals/doctors are already working with AI to make diagnoses. (True)
	15a***	KI kann die Überlebenschance bei Krebspatienten vorhersagen. (Richtig)	AI can predict the chance of survival in cancer patients. (True)
	16a***	Pflegeroboter unterstützen bereits pflegebedürftige Personen mit Aufgaben im Haushalt und der Einnahme ihrer Medikamente. (Richtig)	Care robots already assist people in need of care with household chores and with taking their medications. (True)

	17a*	Eine länderübergreifende Datensammlung von PatientInnen als Informationsquelle für KI ist nicht zu empfehlen, da die Behandlungsmethoden und die Art der Krankheiten zwischen den Ländern zu unterschiedlich sind. (Falsch)	Cross-national data collection from patients as a source of information for AI is not recommended because treatment methods and the nature of diseases vary too much between countries. (False)
	18a*	Eine Diagnose durch eine Künstliche Intelligenz hat den Vorteil, dass der Vorgang vollkommen transparent ist und von den PatientInnen jederzeit einzusehen ist. (Falsch)	A diagnosis by an artificial intelligence has the advantage that the process is completely transparent and can be viewed by the patient at any time. (False)
	19a***	Durch das Scannen von Social-Media-Beiträgen kann eine KI zukünftige Depressionen bei Nutzern vorhersagen. (Richtig)	By scanning social media posts, an AI can predict future depression in users. (True)
	8b**	KI kann noch nicht bei der Entwicklung von Medikamenten helfen. (Falsch)	AI cannot yet help with drug development. (False)
	9b**	KI ist noch nicht in der Lage einzelne Krankheitssymptome zu einer Diagnose korrekt zusammenfügen. (Falsch)	AI is not yet able to correctly combine individual disease symptoms into a diagnosis. (False)
Autonomous Driving	20a*	Vollständig autonome Fahrzeuge, d.h. Fahrzeuge ohne menschlichen Fahrer, sind bereits für den Straßenverkehr in Deutschland zugelassen. (Falsch)	Fully autonomous vehicles, i.e., vehicles without a human driver, are already approved for road traffic in Germany. (False)
	21a***	Ein autonomes Fahrzeug ist ein Fahrzeug, dass von einer KI gesteuert wird. (Richtig)	An autonomous vehicle is a vehicle that is controlled by an AI. (True)
	22a	Ein autonomes Fahrzeug kann nicht von einem Menschen gesteuert werden. (Falsch)	An autonomous vehicle cannot be controlled by a human. (False)
	23a***	Eines der größten Probleme beim Trainieren der KI für das autonome Fahren ist es, Objekte im Fahrbereich korrekt zu erkennen. (Richtig)	One of the biggest problems in training AI for autonomous driving is correctly detecting objects in the driving area. (True)
	24a*	Damit ein autonomes Fahrzeug zugelassen werden kann, muss es in einer Vielzahl an Situationen, von schneebedeckten Bergen bis hin zu Wüstenstraßen, geprüft werden. (Falsch)	For an autonomous vehicle to be approved, it must be tested in a variety of situations, from snow-covered mountains to desert roads. (False)
	25a*	Um eine KI für ein autonomes Fahrzeug zu trainieren, werden Daten aus echten Testfahrten verwendet. (Falsch)	To train an AI for an autonomous vehicle, data from real test drives are used. (False)

	10b**	Autonomes Fahren ist trotz KI Unterstützung nicht ohne Fahrer möglich. (Falsch)	Autonomous driving is not possible without a driver, despite AI support. (False)
	11b**	EPIMETHEUS ist ein wissenschaftliches Projekt, das sich zum Ziel gemacht hat, ein vollständig funktionsfähiges autonomes Auto zu entwickeln. (Falsch)	EPIMETHEUS is a scientific project that aims to develop a fully functional autonomous car. (False)
	12b**	Das erste autonome Auto wurde 2005 entwickelt.(Falsch)	The first autonomous car was developed in 2005. (False)
Education	26a***	Beim adaptiven Testen passt eine KI die Schwierigkeit der Aufgaben an die Fähigkeit des Teilnehmenden im Verlauf der Testung an. (Richtig)	In adaptive testing, an AI adjusts the difficulty of tasks to the participant's ability as the test progresses. (True)
	27a***	In manchen Schulen wird der Unterricht bereits vollkommen von KI durchgeführt. (Falsch)	In some schools, teaching is already being conducted entirely by AI. (False)
	28a***	Eine KI, die Kameraaufnahmen von SchülerInnen in Klassenzimmern auswertet, kann feststellen, worauf die SchülerInnen ihre Aufmerksamkeit richten. (Richtig)	An AI that evaluates camera footage of students in classrooms can determine where students are focusing their attention. (True)
History	29a***	Deep Blue war die erste KI, die einen professionellen Schachspieler schlug. (Richtig)	Deep Blue was the first AI to beat a professional chess player. (True)
	30a*	Der erste Computer, der als KI bezeichnet werden kann hieß ELIZA. (Falsch)	The first computer that can be called AI was called ELIZA. (False)
Personnel recruiting	31a***	Einige Unternehmen lassen eine KI selbständig Stelleninterviews durchführen und eine Vorauswahl an Kandidaten treffen. (Richtig)	Some companies let an AI independently conduct job interviews and pre-select candidates. (True)
	32a	Künstliche Intelligenzen sind frei von Verzerrungen bei der Personalauswahl. (Falsch)	Artificial intelligences are free from bias in personnel selection. (False)
	33a***	Eine KI kann bei der Personalauswahl eine größere Menge an Daten mit einbeziehen, als ein Mensch. (Richtig)	An AI can incorporate a larger amount of data than a human when selecting personnel. (True)
Media	34a***	Viele soziale Medien verwenden KI-basierte Algorithmen, die die angezeigten Inhalte personalisiert auswählen. (Richtig)	Many social media use AI-based algorithms that select the content displayed in a personalized way. (True)
	35a***	Unternehmen können über KI Marketinginhalte für soziale Medien generieren und so in Massen verbreiten. (Richtig)	Companies can use AI to generate marketing content for social media that the companies can mass distribute. (True)
	36a	Mittels KI können Videos erstellt werden, die nicht zu unterscheiden sind von	Using AI, videos can be created that are indistinguishable from

		Videos, die von echten Menschen erstellt wurden. (Falsch)	videos created by real people. (False)
	37a***	Mittels eines Deepfakes können Videos von berühmten Persönlichkeiten gefälscht werden. (Richtig)	Using deepfake, videos of famous people can be faked. (True)
	38a*	Netflix veranstaltete 2006 einen Wettbewerb bei dem ein Sprachassistent entworfen werden sollte, der Vorschläge auf Basis der Nutzungsaktivitäten macht. (Falsch)	Netflix held a contest in 2006 to design a voice assistant that makes suggestions based on usage activities. (False)
	39a*	YouTubes Vorschlags-Algorithmus bezieht Daten über den Nutzer, wie Alter, bisheriges Nutzungsverhalten und die Tageszeit am Standort, mit ein. (Falsch)	YouTube's suggestion algorithm incorporates data about the user, such as age, previous usage patterns, and the time of day at the location of the user. (False)
	13b*	YouTubes Vorschlags-Algorithmus bezieht Daten über den Nutzer, wie Alter und die Tageszeit am Standort, mit ein. (Falsch)	YouTube's suggestion algorithm incorporates data about the user, such as age and the time of day at the location of the user. (False)
	40a***	Reist man von Deutschland nach Frankreich, unterscheiden sich die personalisierten Werbeanzeigen im Web an den zwei Standorten. (Richtig)	If one travels from Germany to France, the personalized ads on the web differ in the two locations. (True)
	41a***	Facebook benutzt KI, um aus Social-Media-Posts die Selbstmord-Neigung von NutzerInnen vorherzusagen. (Richtig)	Facebook uses AI to predict users' suicidal tendencies from social media posts. (True)
Communication	42	Chatbots verarbeiten und verstehen Sprache auf die gleiche Art und Weise wie der Mensch. (Falsch)	Chatbots process and understand language in the same way as humans. (False)
	43***	Ein Chatbot kann die Frage "Wird es morgen regnen?" mit einer hohen Wahrscheinlichkeit korrekt beantworten. (Richtig)	A chatbot can correctly answer the question "Will it rain tomorrow?" with a high probability. (True)
	44***	Die Autokorrektur bei Texteingaben ist eine Form der Künstlichen Intelligenz. (Richtig)	Autocorrection for text input is a form of artificial intelligence. (True)
	14b**	Menschen können immer noch problemlos von KI erzeugte Sprache als künstliche Sprache erkennen. (Falsch)	Humans can still easily recognize AI-generated speech as artificial speech. (False)
	15b**	Das Sprachverständnis von KI-Systemen reicht noch nicht an das vom Menschen heran. (Falsch)	The language understanding of AI systems does not yet reach that of humans. (False)
	16b**	Menschen können Fragen zu einem gelesenen Text besser als KI-Systeme beantworten.	Humans can answer questions about a text they read better than AI systems.

		(Falsch)	(False)
	17b**	Spracherkennungssoftwares wie Alexa oder Apple Siri nutzen keine KI. (Falsch)	Voice recognition software like Alexa or Apple Siri do not use AI. (False)
	18b**	KI-basierte Anwendungen (wie z.B. Gesichtserkennung, Spracherkennung) können von allen Menschen unabhängig von Hautfarbe oder Sprache gleich gut genutzt werden. (Falsch)	AI-based applications (such as facial recognition, speech recognition) can be used equally well by all people regardless of skin color or language. (False)
Environment	45a*	KI ist nicht in der Lage Hurricanes vorherzusagen.(Falsch)	AI is not able to predict hurricanes.(False)
	46***	Der CO2 Ausstoß im Verkehr kann mittels KI-unterstützter Lösungen verringert werden. (Richtig)	CO2 emissions in transport can be reduced using AI-assisted solutions. (True)

Note. Marked items (*) were only used in the final analysis of Study 1. Marked items (**) were only used in the final analysis of Study 2. Marked items (***) were used in the final analysis of Study 1 and Study 2. Items marked with ‘a’ were developed for Study 1, while items marked with ‘b’ were developed for Study 2.

Table A2

Risk-benefit balanced scenarios

Nr.	Domain	German	English
Scenario 1*	Medical 1	Sie entdecken einen immer größer werdenden Fleck auf Ihrer Haut. Aus Sorge, dass es Hautkrebs ist, machen Sie ein Bild von dem Fleck und laden es auf eine Website hoch, wo es durch Künstliche Intelligenz analysiert wird. Die KI kommt nach kurzer Wartezeit zu dem Ergebnis, dass kein bösartiger Hautkrebs vorliegt. Die KI teilt Ihnen das mit und sagt, dass Sie sich keine Sorgen machen müssen. Daraufhin entscheiden Sie sich, keinen Arzttermin zu vereinbaren.	You discover a spot on your skin that is getting bigger and bigger. Worried that it is skin cancer, you take a picture of the spot and upload it to a website where it is analyzed by artificial intelligence. After a short wait, the AI concludes that it is not malignant skin cancer. The AI informs you of this and tells you not to worry. As a result, you decide not to make a doctor's appointment.
Scenario 2*	Medical 2	Sie merken, dass sich Ihre Stimmung in den letzten Wochen ohne ersichtlichen Grund immer weiter verschlechtert. Sie befürchten, dass Sie depressiv sind und recherchieren dazu. Dabei stoßen Sie auf	You notice that your mood has been getting worse and worse over the past few weeks for no apparent reason. You fear that you are depressed and do some research on the subject. In the process,

		eine App, die Ihnen als digitale Psychotherapie auf Basis von Künstlicher Intelligenz vorgestellt wird. Sie laden sich diese App herunter und unterhalten sich mit der KI. Sie beschreiben Ihre Stimmungsschwankungen **(und Ihre Überlegung, einen Therapeuten aufzusuchen) und dass Sie zunehmend weniger Freude an Ihren Hobbys haben** und sich nun überlegen einen Therapeuten aufzusuchen. Die Künstliche Intelligenz teilt Ihnen mit, dass Ihre Stimmungsschwankungen noch keine Depression sind. Sie entscheiden sich keinen Therapeuten aufzusuchen.	you come across an app that is presented to you as digital psychotherapy based on artificial intelligence. You download this app and have a conversation with the AI. You describe your **mood swings and (that you are considering seeing a therapist) that you are finding increasingly less enjoyment in your hobbies and are now considering seeing a therapist.** The AI tells you that your mood swings are not yet depression. You decide not to see a therapist
Scenario 3	Medical 3	Sie kaufen sich eine neue Smartwatch, die neben diversen Neuerungen auch Ihren Herzschlag, Ihre Herzfrequenz und Ihren Blutdruck messen kann. Mithilfe von Künstlicher Intelligenz kann sie Sie auch warnen, wenn bei Ihnen ein irregulärer Herzrhythmus auftritt und Sie einen Doktor aufsuchen sollen. Sie selbst fühlen sich fit und gesund und haben erst vor kurzem einen Gesundheitscheck beim Arzt gemacht. Nachdem Sie die Smartwatch für ein paar Tage getragen haben, warnt die KI Sie jedoch, dass Ihr Herzschlag unregelmäßig ist. Sie beschließen einen Doktor aufzusuchen.	You buy a new smartwatch that can measure your heartbeat, heart rate, and blood pressure, in addition to various innovations. With the help of artificial intelligence, it can also warn you when you have an irregular heart rhythm and need to see a doctor. You yourself feel fit and healthy and have just recently had a health check at the doctor. However, after wearing the smartwatch for a few days, the AI warns you that your heartbeat is irregular. You decide to see a doctor.
Scenario 4*	Driving 1	Sie haben sich ein neues Auto gekauft. Auf der Autobahn probieren Sie die durch Künstliche Intelligenz möglich gemachten technischen Neuerungen zum sicheren Fahren aus. Ihr Auto ist nun in der Lage, eine bestimmte Geschwindigkeit beizubehalten, den Sicherheitsabstand zum Vordermann einzuhalten und auch die Fahrbahn zu halten, ohne dass Sie mit dem Lenkrad korrigierend eingreifen müssen. Dann klingelt Ihr Handy auf der Rückbank. (**Sowohl vor als auch hinter Ihnen ist kein Auto zu sehen.) Die Autobahn ist vierspurig und stark befahren.** Die KI-unterstützten Systeme sind noch eingeschaltet, weshalb Sie umdrehen um Ihr Handy von der Rückbank holen.	You have bought a new car. On the highway, you try out the technical innovations made possible by artificial intelligence for safe driving. Your car is now able to maintain a certain speed, keep a safe distance from the car in front and keep to the lane without you having to take corrective action with the steering wheel. Then your cell phone rings in the back seat. **(There is no car in sight both in front of and behind you.) The highway consists of four very busy lanes.** The AI-assisted systems are still switched on, so you turn around to get your cell phone from the back seat.

Scenario 5*	Driving 2	Nach einem langen Arbeitstag sind Sie in Ihrem Auto auf dem Weg nach Hause. Damit Sie sich auf **der kurvigen Landstraße** nur auf das Lenken konzentrieren müssen, schalten Sie Ihr Auto auf halbautomatisches Fahren. Das bedeutet, das Auto hält die vorgegebene Geschwindigkeit und den Sicherheitsabstand zu anderen Autos ein. Da Ihr Tag sehr lang war, sind Sie sehr müde. Sie merken, dass Ihnen immer häufiger die Augen zufallen und Sie in einen Sekundenschlaf verfallen. Dennoch machen Sie keine Fahrpause und fahren die restlichen 45 Minuten Heimweg mit den eingeschalteten KI-Systemen.	After a long day at work, you are in your car on the way home. So that you can concentrate only on steering on the **serpentine road**, you switch your car to semi-automatic driving. This means the car will maintain the set speed and safe distance from other cars. Since your day was very long, you are very tired. You notice that your eyes are falling shut more and more often and you fall into a microsleep. Nevertheless, you do not take a break from driving and drive the remaining 45 minutes home with the AI systems switched on.
Scenario 6	Driving 3	Sie und ihre Kinder sind mit dem Auto auf dem Rückweg von den Ferien. Es ist auch der letzte Tag Ihres Urlaubs und Sie müssen am darauffolgenden Tag wieder arbeiten. Bezüglich Ihres Arbeitsbeginns erwarten Sie eine wichtige Nachricht von Ihrer Chefin. Ihr Handy haben Sie daher mit dem Autocomputer verbunden. Während der Fahrt erhalten Sie die Information, dass Ihre Chefin Ihnen geschrieben hat. Es dauert jedoch noch, bis Sie zuhause sind. Sie wissen allerdings auch, dass Ihre Chefin zeitnah Feierabend machen wird. Damit Sie die Nachricht Ihrer Chefin lesen und ihr antworten können, schalten Sie den KI-unterstützten Abstand- sowie Fahrbahnlinienhalter Ihres Autos ein und holen Ihr Handy aus der Tasche.	You and your children are on your way back from vacation by car. It is also the last day of your vacation, and you must go back to work the following day. You are expecting an important message from your boss regarding your start of work. Therefore, you have connected your cell phone to the car computer. While driving, you receive the information that your boss has written to you. However, it is still some time before you get home. However, you also know that your boss will be leaving work soon. So that you can read your boss's message and reply to her, you switch on your car's AI-assisted distance and lane line holder and take your cell phone out of your pocket.
Scenario 7*	Media 1	Sie sitzen mit Ihrem Kind im Zug auf dem Weg in den Urlaub. Da Sie wussten, dass die Fahrt länger dauern wird, haben Sie auf Ihrem Tablet ein Spiel aus dem App-Store für Ihr Kind heruntergeladen. Es hat Sie zwar nichts gekostet, dafür müssen Sie im Spiel Werbung in Kauf nehmen. Als Sie nun das Spiel auf dem Tablet öffnen, werden Sie nach diversen Berechtigungen gefragt, die Sie zulassen, damit Ihr Kind das Spiel spielen kann. Dazu gehört auch	You are on the train with your child on the way to vacation. Knowing that the journey will take longer, you have downloaded a game from the app store for your child on your tablet. It did not cost you anything, but you must put up with advertising in the game. When you now open the game on the tablet, you are asked for various permissions that you allow so that your child can play the game. This includes specific data sharing

		eine spezifische Datenfreigabe, um personalisierte Werbung einsetzen zu können. Sie geben Ihrem Kind das Tablet und es beginnt zu spielen. Nach einer Weile hören Sie Geräusche von Schusswaffen, die von dem Tablet kommen und sehen, dass Ihr Kind Werbung für gewalthaltige Videospiele angezeigt bekommt. Sie entscheiden sich dennoch, das kostenfreie Angebot zu nutzen.	to be able to use personalized advertising. You hand the tablet to your child, and he starts playing. After a while, you hear sounds of guns coming from the tablet and see your child being shown ads for violent video games. You still decide to use the free offer.
Scenario 8*	Media 2	Um mit Ihren Freunden in Kontakt zu bleiben, sind Sie regelmäßig auf sozialen Netzwerken unterwegs. Wenn Ihnen etwas gefällt, markieren Sie dies und manchmal teilen Sie auch den Inhalt. Außerdem posten Sie selbst ab und an etwas und stellen Bilder online. (**In letzter Zeit fühlen Sie sich etwas niedergeschlagen.)** Auch heute posten Sie ein Bild von sich, auf dem Sie lächeln und die Sonne scheint und schreiben einen kurzen Text dazu: „A good day to die!". Kurz darauf erhalten Sie eine Nachricht von einem Facebook-Mitarbeiter, dass eine Künstliche Intelligenz durch die Analyse Ihrer vergangenen Posts auf Sie aufmerksam geworden ist und die Mitarbeiter bei Facebook vor suizidalen Absichten bei Ihnen gewarnt hat. Daher legt der Mitarbeiter Ihnen in seiner Nachricht Beratungsstellen ans Herzen und gibt Ihnen Tipps zum Umgang mit Depressionen. Obwohl Sie denken, dass Ihr Post ironisch gemeint war, sind Sie doch so verunsichert, dass Sie nach einer Beratungsstelle googeln.	To stay in touch with your friends, you are regularly on social networks. When you like something, you tag it and sometimes you share the content. You also post something yourself from time to time and put pictures online. **(Lately, you have been feeling a little down.)** Also, today you post a picture of yourself smiling and the sun shining and write a short text to it: "A good day to die!". Shortly thereafter, you receive a message from a Facebook employee stating that an artificial intelligence has become aware of you by analyzing your past posts and has warned the employees at Facebook about suicidal intentions on your part. Therefore, in his message, the employee suggests counseling services to you and gives you tips on how to deal with depression. Although you think that your post was meant ironically, you are so unsettled that you Google for a counseling center.
Scenario 9	Media 3	Sie und Ihre Freunde haben eine neue App entdeckt. Diese App kann mit Hilfe von Künstlicher Intelligenz in kurzen Videoclips das Gesicht eines Prominenten durch ein anderes Gesicht austauschen. Dazu muss man lediglich ein Bild z.B. von sich selbst in der App hochladen und das entsprechende Wunschvideo auswählen. In dem Video sieht es dann so aus, als ob man selbst die Wörter sagt und die Bewegungen	You and your friends have discovered a new app. This app can use artificial intelligence to replace the face of a celebrity with another face in short videos. All you must do is upload a picture of yourself, for example, to the app and select the corresponding video of your choice. In the video, it then looks as if you yourself are saying the words and making the movements that the celebrity

	macht, die der Prominente in dem Video gemacht hat. Die Qualität der Videos ist dabei so gut, dass man kaum merkt, dass das Gesicht verändert wurde. Sie und Ihre Freunde haben damit bereits diverse lustige Videos mit eigenen Bildern erstellt. Als Sie wieder mal ein Video erstellen wollen, stellen Sie fest, dass es inzwischen auch die Möglichkeit gibt, Gesichter in pornografischen Videoclips zu transformieren. Sie beschließen die App dennoch weiterhin zu nutzen.	did in the video. The quality of the videos is so good that you hardly notice that the face has been changed. You and your friends have already used it to create various funny videos with your own pictures. When you want to create a video again, you realize that there is now also the possibility to transform faces into pornographic video clips. You decide to continue using the app anyway.

Note. Marked scenarios (*) were used in the analysis of Study 2. Sentences that were changed for Study 2 are displayed in bold font, the original sentences (Study 1) are displayed in bold font and in brackets.

Table A3

Risk-benefit assessment

Category	Risk/Benefit	German	Answer options	English	Answer options
Affect	Risk	Wie **besorgt**, wenn überhaupt, sind Sie, wenn Sie an diesen Einsatz von KI denken?	überhaupt nicht, leicht, mäßig, sehr, extrem **besorgt**	How concerned, if at all, are you when you think about this use of AI?	not at all, slightly, moderately, very, extremely **worried**
Affect	Risk	Wenn Sie an diesen Einsatz von KI denken, in welchem Ausmaß sind Sie **beunruhigt**?	überhaupt nicht, leicht, mäßig, sehr, extrem **beunruhigt**	When you think about this use of AI, to what extent are you **troubled**?	not at all, slightly, moderately, very, extremely **troubled**
Affect	Risk	Wie **pessimistisch**, wenn überhaupt, sind Sie, wenn Sie sich diesen Einsatz von KI vorstellen?	überhaupt nicht, leicht, mäßig, sehr, extrem **pessimistisch**	How **pessimistic**, if at all, are you about this use of AI?	not at all, slightly, moderately, very, extremely **pessimistic**
Susceptibility	Risk	Wenn Sie diesen Einsatz von KI in Ihrem Umfeld erleben	überhaupt nicht, einigermaßen, mittel, sehr,	If you were to experience this use of AI in your	not at all, somewhat, moderately, very, extremely **likely**

		würden, wie **wahrscheinlich** würde es Sie negativ beeinflussen?	extrem **wahrscheinlich**	environment, how **likely** would it negatively impact you?	
Susceptibility	Risk	Wenn Sie diesen Einsatz von KI in Ihrem Umfeld erleben würden, wie **betroffen** wären Sie von den Konsequenzen?	überhaupt nicht, leicht, mäßig, sehr, extrem **betroffen**	If you were to experience this use of AI in your environment, how **affected** would you be by the consequences?	not at all, slightly, moderately, very, extremely **affected**
Susceptibility	Risk	Wenn Sie diesen Einsatz von KI in Ihrem Umfeld erleben würden, wie **wahrscheinlich** ist es, dass Sie darunter leiden würden?	überhaupt nicht, einigermaßen, mittel, sehr, extrem **wahrscheinlich**	If you were to experience this use of AI in your environment, how **likely** is it that you would suffer?	not at all, somewhat, moderately, very, extremely **likely**
Affect	Benefit	Wie **optimistisch**, wenn überhaupt, sind Sie, wenn Sie an diesen Einsatz von KI denken?	überhaupt nicht, leicht, mäßig, sehr, extrem **optimistisch**	How **optimistic**, if at all, are you when you think about this use of AI?	not at all, slightly, moderately, very, extremely **optimistic**
Affect	Benefit	Wenn Sie an diesen Einsatz von KI denken, in welchem Ausmaß sind Sie **beruhigt**?	überhaupt nicht, leicht, mäßig, sehr, extrem **beruhigt**	When you think about this use of AI, to what extent are you **reassured**?	not at all, slightly, moderately, very, extremely **reassured**
Affect	Benefit	Wie **zuversichtlich**, wenn überhaupt, sind Sie, wenn Sie sich diesen Einsatz von KI vorstellen?	überhaupt nicht, leicht, mäßig, sehr, extrem **zuversichtlich**	How **confident**, if at all, are you in imagining this use of AI?	not at all, slightly, moderately, very, extremely **confident**
Susceptibility	Benefit	Wenn Sie diesen Einsatz von KI in Ihrem Umfeld erleben würden, wie **wahrscheinlich** würde es Sie positiv beeinflussen?	überhaupt nicht, einigermaßen, mittel, sehr, extrem **wahrscheinlich**	If you were to experience this use of AI in your environment, how **likely** would it positively impact you?	not at all, slightly, moderately, very, extremely **likely**

Susceptibility	Benefit	Wenn Sie diesen Einsatz von KI in Ihrem Umfeld erleben würden, wie sehr würden sie von den Konsequenzen **profitieren**?	überhaupt nicht, leicht, mäßig, sehr, extrem **profitieren**	If you were to see this use of AI in your environment, how **beneficial** would the consequences be for you?	not at all, slightly, moderately, very, extremely **beneficial**
Susceptibility	Benefit	Wenn Sie diesen Einsatz von KI in Ihrem Umfeld erleben würden: wie **wahrscheinlich** ist es, dass Sie davon profitieren würden?	überhaupt nicht, einigermaßen, mittel, sehr, extrem **wahrscheinlich**	If you were to see this use of AI in your environment, how **likely** is it that you would benefit?	not at all, slightly, moderately, very, extremely **likely**

Note. Items were adapted from Walpole et al. (2021)

Appendix B

Regression results Study 1

Table B1

Regression results using RB1 score

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	1.02**	[0.51, 1.52]				
Education	0.02	[-0.09, 0.13]	.00	[-.00, .00]		
Age	-0.03**	[-0.04, -0.01]	.03	[-.00, .06]		
Gender	-0.22	[-0.50, 0.05]	.01	[-.01, .02]		
					R^2 = .042** 95% CI[.01,.08]	
(Intercept)	0.37	[-0.44, 1.18]				
Education	0.00	[-0.11, 0.11]	.00	[-.00, .00]		
Age	-0.03**	[-0.04, -0.01]	.03	[-.00, .06]		
Gender	-0.14	[-0.43, 0.14]	.00	[-.01, .01]		
Confidence	0.19*	[0.00, 0.38]	.01	[-.01, .03]		
					R^2 = .052** 95% CI[.01,.09]	ΔR^2 = .010* 95% CI[-.01, .03]
(Intercept)	0.16	[-1.09, 1.41]				
Education	0.00	[-0.11, 0.11]	.00	[-.00, .00]		

Age	-0.03**	[-0.04, -0.01]	.03	[-.00, .07]		
Gender	-0.14	[-0.43, 0.15]	.00	[-.01, .01]		
Confidence	0.19	[-0.01, 0.38]	.01	[-.01, .03]		
Knowledge	0.38	[-1.32, 2.09]	.00	[-.00, .00]		
					R^2 = .053** 95% CI[.01,.09]	ΔR^2 = .000 95% CI[-.00, .00]

Note. A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05. ** indicates p < .01.

Table B2

Regression results using RB2 score

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	0.83**	[0.52, 1.15]				
Education	0.06	[-0.01, 0.13]	.01	[-.01, .03]		
Age	-0.00	[-0.01, 0.01]	.00	[-.00, .00]		
Gender	0.12	[-0.05, 0.30]	.01	[-.01, .02]		
					R^2 = .016 95% CI[.00,.04]	
(Intercept)	-0.33	[-1.05, 0.38]				
Education	0.06	[-0.01, 0.13]	.01	[-.01, .02]		
Age	-0.00	[-0.01, 0.01]	.00	[-.01, .01]		
Gender	0.16	[-0.01, 0.33]	.01	[-.01, .03]		
Knowledge	1.88**	[0.84, 2.91]	.03	[-.00, .07]		
					R^2 = .047** 95% CI[.01,.09]	ΔR^2 = .031** 95% CI[-.00, .07]
(Intercept)	-0.71	[-1.48, 0.06]				
Education	0.05	[-0.02, 0.12]	.00	[-.01, .02]		
Age	-0.00	[-0.01, 0.01]	.00	[-.00, .01]		
Gender	0.22*	[0.04, 0.40]	.01	[-.01, .04]		
Knowledge	1.65**	[0.60, 2.69]	.02	[-.01, .05]		
Confidence	0.15*	[0.04, 0.27]	.02	[-.01, .04]		
					R^2 = .063** 95% CI[.02,.10]	ΔR^2 = .016* 95% CI[-.01, .04]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table B3

Regression results using benefit as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	3.24**	[2.93, 3.56]				
Education	0.01	[-0.06, 0.08]	.00	[-.00, .00]		
Age	-0.02**	[-0.03, -0.01]	.04	[-.00, .07]		
Gender	-0.08	[-0.25, 0.09]	.00	[-.01, .01]		
					R^2 = .043** 95% CI[.01,.08]	
(Intercept)	2.62**	[2.12, 3.12]				
Education	-0.01	[-0.08, 0.06]	.00	[-.00, .00]		
Age	-0.02**	[-0.03, -0.01]	.04	[-.00, .07]		
Gender	-0.00	[-0.18, 0.17]	.00	[-.00, .00]		
Confidence	0.19**	[0.07, 0.30]	.02	[-.01, .05]		
					R^2 = .066** 95% CI[.02,.11]	ΔR^2 = .023** 95% CI[-.01, .05]
(Intercept)	2.99**	[2.22, 3.76]				
Education	-0.01	[-0.08, 0.06]	.00	[-.00, .00]		
Age	-0.02**	[-0.03, -0.01]	.03	[-.00, .07]		
Gender	-0.01	[-0.19, 0.17]	.00	[-.00, .00]		
Confidence	0.20**	[0.08, 0.32]	.03	[-.00, .06]		
Knowledge	-0.66	[-1.72, 0.39]	.00	[-.01, .02]		
					R^2 = .070** 95% CI[.02,.11]	ΔR^2 = .004 95% CI[-.01, .02]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL*

indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table B4

Regression results using risk as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	2.23**	[1.94, 2.51]				
Education	-0.01	[-0.07, 0.06]	.00	[-.00, .00]		
Age	0.01*	[0.00, 0.02]	.01	[-.01, .03]		
Gender	0.15	[-0.01, 0.30]	.01	[-.01, .03]		
					R^2 = .022* 95% CI[.00,.05]	
(Intercept)	2.25**	[1.78, 2.71]				
Education	-0.01	[-0.07, 0.06]	.00	[-.00, .00]		
Age	0.01*	[0.00, 0.02]	.01	[-.01, .03]		
Gender	0.14	[-0.02, 0.31]	.01	[-.01, .02]		
Confidence	-0.01	[-0.12, 0.10]	.00	[-.00, .00]		
					R^2 = .022 95% CI[.00,.05]	ΔR^2 = .000 95% CI[-.00, .00]
(Intercept)	2.83**	[2.12, 3.54]				
Education	-0.01	[-0.07, 0.06]	.00	[-.00, .00]		
Age	0.01*	[0.00, 0.02]	.01	[-.01, .04]		
Gender	0.13	[-0.03, 0.29]	.01	[-.01, .02]		
Confidence	0.01	[-0.10, 0.12]	.00	[-.00, .00]		
Knowledge	-1.05*	[-2.02, -0.08]	.01	[-.01, .03]		
					R^2 = .033* 95% CI[.00,.06]	ΔR^2 = .011* 95% CI[-.01, .03]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Appendix C

Regression results Study 2

Table C1

Regression results using RB1 score

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	0.13	[-0.47, 0.73]				
Education	0.06	[-0.04, 0.17]	.00	[-.01, .01]		
Age	-0.01	[-0.02, 0.00]	.01	[-.01, .02]		
Gender	-0.43**	[-0.68, -0.17]	.02	[-.00, .05]		
					R^2 = .036** 95% CI[.01,.07]	
(Intercept)	-0.24	[-0.91, 0.44]				
Education	0.04	[-0.06, 0.15]	.00	[-.01, .01]		
Age	-0.01	[-0.02, 0.00]	.01	[-.01, .02]		
Gender	-0.34*	[-0.60, -0.08]	.01	[-.01, .04]		
Confidence	0.14*	[0.02, 0.25]	.01	[-.01, .03]		
					R^2 = .048** 95% CI[.01,.09]	ΔR^2 = .012* 95% CI[-.01, .03]
(Intercept)	-1.10*	[-1.95, -0.25]				
Education	0.01	[-0.09, 0.12]	.00	[-.00, .00]		
Age	-0.01*	[-0.02, -0.00]	.01	[-.01, .03]		
Gender	-0.34*	[-0.60, -0.08]	.01	[-.01, .04]		
Confidence	0.08	[-0.04, 0.20]	.00	[-.01, .02]		
Knowledge	1.96**	[0.75, 3.16]	.02	[-.00, .05]		
					R^2 = .070** 95% CI[.02,.11]	ΔR^2 = .022** 95% CI[-.00, .05]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table C2

Regression results using the RB2 score

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	0.30	[-0.13, 0.72]				
Education	0.06	[-0.01, 0.13]	.01	[-.01, .02]		
Age	0.01**	[0.00, 0.02]	.03	[-.00, .06]		
Gender	0.00	[-0.18, 0.18]	.00	[-.00, .00]		
					R^2 = .032** 95% CI[.00,.07]	
(Intercept)	0.05	[-0.55, 0.65]				
Education	0.05	[-0.03, 0.12]	.00	[-.01, .02]		
Age	0.01**	[0.00, 0.02]	.03	[-.00, .06]		
Gender	0.01	[-0.17, 0.19]	.00	[-.00, .00]		
Knowledge	0.48	[-0.34, 1.31]	.00	[-.01, .01]		
					R^2 = .035** 95% CI[.00,.07]	ΔR^2 = .003 95% CI[-.01, .01]
(Intercept)	-0.09	[-0.70, 0.51]				
Education	0.04	[-0.04, 0.11]	.00	[-.01, .01]		
Age	0.01**	[0.00, 0.02]	.03	[-.00, .06]		
Gender	0.08	[-0.10, 0.27]	.00	[-.01, .01]		
Knowledge	0.12	[-0.73, 0.98]	.00	[-.00, .00]		
Confidence	0.12**	[0.04, 0.21]	.02	[-.01, .04]		
					R^2 = .053** 95% CI[.01,.09]	ΔR^2 = .018** 95% CI[-.01, .04]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table C3

Regression results using benefit as the criterion

Predictor	*b*	*b* 95% CI	sr^2	sr^2 95% CI	Fit	Difference

		[LL, UL]		[LL, UL]		
(Intercept)	3.22**	[2.84, 3.61]				
Education	-0.02	[-0.09, 0.04]	.00	[-.01, .01]		
Age	-0.01**	[-0.02, -0.01]	.04	[.00, .07]		
Gender	-0.22**	[-0.38, -0.06]	.02	[-.01, .04]		
					R^2 = .047**	
					95% CI[.01,.09]	
(Intercept)	3.00**	[2.57, 3.43]				
Education	-0.04	[-0.10, 0.03]	.00	[-.01, .01]		
Age	-0.01**	[-0.02, -0.01]	.04	[.00, .07]		
Gender	-0.17*	[-0.34, -0.00]	.01	[-.01, .03]		
Confidence	0.08*	[0.01, 0.16]	.01	[-.01, .03]		
					R^2 = .058**	ΔR^2 = .011*
					95% CI[.02,.10]	95% CI[-.01, .03]
(Intercept)	2.88**	[2.33, 3.43]				
Education	-0.04	[-0.11, 0.03]	.00	[-.01, .01]		
Age	-0.01**	[-0.02, -0.01]	.04	[.00, .07]		
Gender	-0.17*	[-0.34, -0.00]	.01	[-.01, .03]		
Confidence	0.08	[-0.00, 0.15]	.01	[-.01, .02]		
Knowledge	0.27	[-0.50, 1.05]	.00	[-.00, .01]		
					R^2 = .059**	ΔR^2 = .001
					95% CI[.02,.10]	95% CI[-.00, .01]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table C4

Regression results using risk as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	3.09**	[2.69, 3.49]				
Education	-0.09*	[-0.16, -0.02]	.01	[-.01, .04]		
Age	-0.00	[-0.01, 0.00]	.00	[-.01, .01]		
Gender	0.20*	[0.04, 0.37]	.01	[-.01, .03]		
					R^2 = .033**	

					95% CI[.00,.07]	
(Intercept)	3.24**	[2.79, 3.68]				
Education	-0.08*	[-0.15, -0.01]	.01	[-.01, .03]		
Age	-0.00	[-0.01, 0.00]	.00	[-.01, .01]		
Gender	0.17	[-0.00, 0.34]	.01	[-.01, .02]		
Confidence	-0.05	[-0.13, 0.02]	.00	[-.01, .02]		
					R^2 = .038** 95% CI[.01,.07]	ΔR^2 = .004 95% CI[-.01, .02]
(Intercept)	3.98**	[3.42, 4.54]				
Education	-0.06	[-0.13, 0.01]	.01	[-.01, .02]		
Age	-0.00	[-0.01, 0.00]	.00	[-.01, .01]		
Gender	0.17	[-0.00, 0.34]	.01	[-.01, .02]		
Confidence	-0.01	[-0.09, 0.07]	.00	[-.00, .00]		
Knowledge	-1.68**	[-2.47, -0.89]	.04	[.00, .07]		
					R^2 = .076** 95% CI[.03,.12]	ΔR^2 = .038** 95% CI[.00, .07]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table C5

Regression results using RB1 score as the criterion, adding attitudes towards AI

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	2.07**	[1.29, 2.85]				
Education	-0.03	[-0.12, 0.07]	.00	[-.00, .00]		
Age	-0.01*	[-0.02, -0.00]	.01	[-.01, .03]		
Gender	-0.26*	[-0.50, -0.02]	.01	[-.01, .02]		
AI Attitudes	0.30**	[0.24, 0.36]	.17	[.10, .23]		
Confidence	0.02	[-0.09, 0.13]	.00	[-.00, .00]		
					R^2 = .213** 95% CI[.14,.27]	
(Intercept)	1.50**	[0.54, 2.46]				
Education	-0.04	[-0.14, 0.06]	.00	[-.00, .01]		
Age	-0.01*	[-0.02, -0.00]	.01	[-.01, .03]		

Gender	-0.26*	[-0.50, -0.03]	.01	[-.01, .02]		
AI Attitudes	0.29**	[0.23, 0.35]	.15	[.09, .21]		
Confidence	-0.01	[-0.12, 0.10]	.00	[-.00, .00]		
Knowledge	1.12*	[0.01, 2.24]	.01	[-.01, .02]		
					R^2 = .220** 95% CI[.15,.28]	ΔR^2 = .007* 95% CI[-.01, .02]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table C6

Regression results using the RB2 score as the criterion, adding attitudes towards AI

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	-0.18	[-0.91, 0.55]				
Education	0.06	[-0.02, 0.13]	.00	[-.01, .02]		
Age	0.01**	[0.00, 0.02]	.03	[-.00, .06]		
Gender	0.00	[-0.18, 0.18]	.00	[-.00, .00]		
AI Attitudes	-0.03	[-0.07, 0.02]	.00	[-.01, .01]		
Knowledge	0.59	[-0.25, 1.43]	.00	[-.01, .02]		
					R^2 = .037** 95% CI[.00,.07]	
(Intercept)	-0.46	[-1.20, 0.28]				
Education	0.04	[-0.03, 0.12]	.00	[-.01, .01]		
Age	0.01**	[0.01, 0.02]	.03	[-.00, .06]		
Gender	0.07	[-0.11, 0.26]	.00	[-.01, .01]		
AI Attitudes	-0.04	[-0.09, 0.01]	.01	[-.01, .02]		
Knowledge	0.24	[-0.62, 1.10]	.00	[-.00, .01]		
Confidence	0.14**	[0.05, 0.22]	.02	[-.00, .05]		
					R^2 = .059** 95% CI[.01,.09]	ΔR^2 = .022** 95% CI[-.00, .05]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL*

indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table C7

Regression results using benefit as the criterion, adding attitudes towards AI

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	4.11**	[3.65, 4.56]				
Education	-0.06	[-0.13, 0.00]	.01	[-.01, .02]		
Age	-0.01**	[-0.02, -0.01]	.04	[.01, .08]		
Gender	-0.15	[-0.31, 0.00]	.01	[-.01, .02]		
AI Attitudes	0.13**	[0.09, 0.17]	.08	[.03, .13]		
					R^2 = .130** 95% CI[.07,.18]	
(Intercept)	4.04**	[3.39, 4.69]				
Education	-0.07*	[-0.13, -0.00]	.01	[-.01, .02]		
Age	-0.01**	[-0.02, -0.01]	.04	[.01, .07]		
Gender	-0.14	[-0.30, 0.03]	.01	[-.01, .02]		
AI Attitudes	0.13**	[0.09, 0.17]	.07	[.03, .12]		
Knowledge	-0.10	[-0.85, 0.66]	.00	[-.00, .00]		
Confidence	0.04	[-0.04, 0.11]	.00	[-.01, .01]		
					R^2 = .132** 95% CI[.07,.18]	ΔR^2 = .002 95% CI[-.01, .01]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table C8

Regression results using risk as the criterion, adding attitudes towards AI

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference

(Intercept)	1.97**	[1.51, 2.43]				
Education	-0.04	[-0.10, 0.03]	.00	[-.01, .01]		
Age	-0.00	[-0.01, 0.00]	.00	[-.01, .01]		
Gender	0.12	[-0.04, 0.28]	.00	[-.01, .02]		
AI Attitudes	-0.17**	[-0.21, -0.13]	.13	[.07, .19]		
					R^2 = .162**	
					95% CI[.10,.22]	
(Intercept)	2.54**	[1.90, 3.19]				
Education	-0.02	[-0.09, 0.04]	.00	[-.00, .01]		
Age	-0.00	[-0.01, 0.00]	.00	[-.00, .01]		
Gender	0.13	[-0.03, 0.29]	.00	[-.01, .02]		
AI Attitudes	-0.16**	[-0.20, -0.12]	.11	[.05, .16]		
Knowledge	-1.22**	[-1.97, -0.47]	.02	[-.00, .04]		
Confidence	0.04	[-0.03, 0.12]	.00	[-.01, .01]		
					R^2 = .181**	ΔR^2 = .020**
					95% CI[.11,.24]	95% CI[-.00, .04]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Appendix D

Influence of metacognition

Table D1

Regression results using the RB2 score (Study 1)

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	0.83**	[0.52, 1.15]				
Education	0.06	[-0.01, 0.13]	.01	[-.01, .03]		
Age	-0.00	[-0.01, 0.01]	.00	[-.00, .00]		
Gender	0.12	[-0.05, 0.30]	.01	[-.01, .02]		
					R^2 = .016	
					95% CI[.00,.04]	
(Intercept)	0.73**	[0.42, 1.05]				

Education	0.06	[-0.01, 0.13]	.01	[-.01, .02]		
Age	-0.00	[-0.01, 0.01]	.00	[-.00, .00]		
Gender	0.13	[-0.04, 0.30]	.01	[-.01, .02]		
meta-d'	0.30**	[0.12, 0.47]	.03	[-.00, .06]		
					R^2 = .043** 95% CI[.01,.08]	ΔR^2 = .027** 95% CI[-.00, .06]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table D2

Regression results using the RB2 score (Study 2)

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	0.17	[-0.36, 0.70]				
Education	0.06	[-0.01, 0.14]	.01	[-.01, .02]		
Age	0.01**	[0.00, 0.02]	.03	[-.00, .06]		
Gender	-0.01	[-0.19, 0.17]	.00	[-.00, .00]		
AI Attitudes	-0.02	[-0.07, 0.03]	.00	[-.01, .01]		
					R^2 = .033** 95% CI[.00,.06]	
(Intercept)	0.11	[-0.41, 0.63]				
Education	0.04	[-0.03, 0.11]	.00	[-.01, .01]		
Age	0.01**	[0.00, 0.02]	.03	[-.00, .05]		
Gender	0.04	[-0.14, 0.22]	.00	[-.00, .00]		
AI Attitudes	-0.03	[-0.08, 0.02]	.00	[-.01, .01]		
meta-d'	0.26**	[0.11, 0.40]	.03	[-.00, .06]		
					R^2 = .061** 95% CI[.02,.10]	ΔR^2 = .028** 95% CI[-.00, .06]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL*

indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Table D3

Regression results using the benefit score as the criterion, with meta-d' and d-prime as interaction term (Study 2)

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit	Difference
(Intercept)	3.00**	[2.57, 3.43]				
Education	-0.04	[-0.10, 0.03]	.00	[-.01, .01]		
Age	-0.01**	[-0.02, -0.01]	.04	[.00, .07]		
Gender	-0.17*	[-0.34, -0.00]	.01	[-.01, .03]		
Confidence	0.08*	[0.01, 0.16]	.01	[-.01, .03]		
					R^2 = .058** 95% CI[.02,.10]	
(Intercept)	2.97**	[2.50, 3.43]				
Education	-0.04	[-0.10, 0.03]	.00	[-.01, .01]		
Age	-0.01**	[-0.02, -0.01]	.03	[.00, .07]		
Gender	-0.18*	[-0.35, -0.01]	.01	[-.01, .03]		
Confidence	0.08*	[0.00, 0.17]	.01	[-.01, .02]		
metad_c	-0.18*	[-0.32, -0.03]	.01	[-.01, .03]		
dprime_c	0.06	[-0.08, 0.19]	.00	[-.01, .01]		
metad_c:dprime_c	0.17*	[0.02, 0.33]	.01	[-.01, .03]		
					R^2 = .079** 95% CI[.03,.12]	ΔR^2 = .021* 95% CI[-.00, .05]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Appendix E

Table E1

Regression results using RB1 calculated as the difference between the risk and benefit scores, adding attitudes towards AI as moderator

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit
(Intercept)	-0.16	[-0.96, 0.64]			
Education	-0.04	[-0.14, 0.05]	.00	[-.00, .01]	
Age	-0.01**	[-0.02, -0.00]	.01	[-.01, .03]	
Gender	-0.28*	[-0.51, -0.04]	.01	[-.01, .03]	
AI Attitudes_c	0.28**	[0.22, 0.34]	.14	[.08, .20]	
Confidence_c	-0.01	[-0.12, 0.10]	.00	[-.00, .00]	
Knowledge	1.19*	[0.07, 2.30]	.01	[-.01, .02]	
AI Attitudes_c:Confidence_c	0.05*	[0.00, 0.10]	.01	[-.01, .02]	
					R^2 = .228** 95% CI[.15,.28]

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.